

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 20:50:21 ; Search time 16939.5 Seconds  
(without alignments)  
11655.272 Million cell updates/sec

Title: US-10-017-410-1

Perfect score: 4175

Sequence: 1 ggacaggggtgtctcgat.....aaaaaaaaaaaaaaaaaac 4175

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Query Match % | Length | DB ID | Description        |
|------------|--------|---------------|--------|-------|--------------------|
| 1          | 4174   | 100.0         | 4174   | 10    | AF282864 Mus muscu |
| 2          | 3864.4 | 92.6          | 4078   | 10    | BC059819 Mus muscu |
| 3          | 3467.6 | 83.1          | 90370  | 10    | BC005084 Mouse DNA |
| 4          | 3467.6 | 83.1          | 207360 | 2     | BX324222 Mus muscu |
| 5          | 1222.4 | 29.3          | 184163 | 2     | AC120245 Rattus no |
| 6          | 1222.4 | 29.3          | 249734 | 2     | AC097362 Rattus no |
| 7          | 722.2  | 17.3          | 2852   | 6     | C0842051 Sequence  |
| 8          | 722.2  | 17.3          | 2852   | 9     | AK123581 Homo sapi |
| 9          | 705    | 16.9          | 149462 | 9     | AC017081 Homo sapi |
| 10         | 653.8  | 15.7          | 822    | 6     | AX206799 Sequence  |
| 11         | 510.8  | 12.2          | 531    | 11    | G76816 S208P6606FG |
| 12         | 400.4  | 9.6           | 595    | 6     | AX400443 Sequence  |
| 13         | 382.6  | 9.2           | 784    | 11    | BV053534 S212P6219 |
| 14         | 261.2  | 6.3           | 315    | 6     | CQ736905 Sequence  |
| 15         | 261.2  | 6.3           | 315    | 6     | CQ736906 Sequence  |
| 16         | 258    | 6.2           | 315    | 6     | CQ744096 Sequence  |
| 17         | 232.4  | 5.6           | 2318   | 3     | AY071232 Drosophil |
| 18         | 229.2  | 5.5           | 1163   | 3     | AF323976 Drosophil |
| 19         | 205    | 4.9           | 1429   | 3     | AK116177 Ciona int |

|    |       |     |        |    |          |                    |
|----|-------|-----|--------|----|----------|--------------------|
| 20 | 183.6 | 4.4 | 163542 | 9  | AL158206 | Human DNA          |
| 21 | 183.6 | 4.4 | 219679 | 2  | AC109451 | AC109451 Homo sapi |
| 22 | 174.8 | 4.2 | 792    | 6  | AX206797 | Sequence           |
| 23 | 174.8 | 4.2 | 795    | 9  | AF347024 | AF347024 Homo sapi |
| 24 | 173.8 | 4.2 | 822    | 10 | AF347023 | AF347023 Mus muscu |
| 25 | 160.6 | 3.8 | 537    | 6  | CQ605547 | CQ605547 Sequence  |
| 26 | 131.2 | 3.1 | 1817   | 9  | AK024898 | AK024898 Homo sapi |
| 27 | 124   | 3.0 | 112713 | 9  | AL391834 | AL391834 Human DNA |
| 28 | 124   | 3.0 | 219679 | 2  | AC109451 | AC109451 Homo sapi |
| 29 | 114.2 | 2.7 | 358    | 6  | AR419744 | AR419744 Sequence  |
| 30 | 114.2 | 2.7 | 358    | 6  | AX980438 | AX980438 Sequence  |
| 31 | 114.2 | 2.7 | 358    | 6  | BD115297 | BD115297 EST and e |
| 32 | 108   | 2.6 | 2143   | 6  | CQ492597 | CQ492597 Sequence  |
| 33 | 108   | 2.6 | 2143   | 6  | CQ492742 | CQ492742 Sequence  |
| 34 | 108   | 2.6 | 2143   | 6  | CQ493050 | CQ493050 Sequence  |
| 35 | 92.6  | 2.2 | 4758   | 6  | CQ605546 | CQ605546 Sequence  |
| 36 | 92.6  | 2.2 | 13715  | 2  | AC017332 | AC017332 Drosophil |
| 37 | 92.6  | 2.2 | 78857  | 3  | AC004364 | AC004364 Drosophil |
| 38 | 92.6  | 2.2 | 157875 | 3  | AE003665 | AE003665 Drosophil |
| 39 | 92.6  | 2.2 | 164361 | 3  | AC009252 | AC009252 Drosophil |
| 40 | 92.6  | 2.2 | 165585 | 3  | AC093044 | AC093044 Drosophil |
| 41 | 78.4  | 1.9 | 257    | 6  | CQ731034 | CQ731034 Sequence  |
| 42 | 77.8  | 1.9 | 4071   | 6  | CQ615131 | CQ615131 Sequence  |
| 43 | 77.8  | 1.9 | 38138  | 2  | AC017406 | AC017406 Drosophil |
| 44 | 75.4  | 1.8 | 184065 | 2  | AC108371 | AC108371 Pan trogl |
| 45 | 72    | 1.7 | 173696 | 10 | AL670603 | AL670603 Mouse DNA |

## ALIGNMENTS

RESULT 1  
AF282864  
LOCUS Mus musculus cancer related gene-liver 1 mRNA ROD 12-DEC-2001  
DEFINITION AF282864 4174 bp mRNA linear complete cds.  
ACCESSION AF282864  
VERSION AF282864.1 GI:17529683

KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Graveel,C.R., Jatkoa,T., Madore,S.J., Holt,A.L. and Farnham,P.J.  
TITLE Expression profiling and identification of novel genes in hepatocellular carcinomas  
JOURNAL Hepatocellular Carcinoma  
MEDLINE Oncogene 20 (21), 2704-2712 (2001)  
PUBMED 21313787  
11420682

REFERENCE  
AUTHORS Graveel,C.R., Jatkoa,T., Madore,S.J., Holt,A.L. and Farnham,P.J.  
TITLE Identification of genes deregulated in murine hepatocellular carcinomas using oligonucleotide microarrays and representational difference analysis  
JOURNAL Unpublished  
MEDLINE 3 (bases 1 to 4174)  
PUBMED 11420682

REFERENCE  
AUTHORS Graveel,C.R., Jatkoa,T., Madore,S.J., Holt,A.L. and Farnham,P.J.  
TITLE Direct Submission  
JOURNAL Submitted (27-JUN-2000) Oncology, University of Wisconsin, 1400 University Avenue, Madison, WI 53706, USA

FEATURES  
source

1. 4174  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/tissue\_type="hepatocellular carcinoma"  
35..862  
/notes="CRG-LI; putative"  
/codon\_start=1  
/product="cancer related gene-liver 1"  
/protein\_id="AAL40408.1"  
/db\_xref="GI:17529684"  
/translation="MGAPHWHDHLAGSSEVDNCDNYTIVPAIEFYNTISNVLFFI"

CDS

LPPIQWCLPROVATCFNSGIYLIWTLVVVGIGSVYFHATLSFLQMLDELAILWVLM  
CALAMFFPRVLPKIFRNDGRBKAIVCVLSAITTCGLAFIKPAINNIIMLILGDPCTA  
LLVAILKCDNRVYFKLGFSLWSTLALFVWISDQAFCELLSSFFHFFVYLHCVHILFI  
CLASLIGCVCFAYFDAASEIPBQGPVIRFWPSEKWAFIGVPYVSLLCALHKGSPVKIT"

## ORIGIN

|                       |     |  |               |           |              |  |  |
|-----------------------|-----|--|---------------|-----------|--------------|--|--|
| Query Match           |     | 100.0%;  | Score 4174;   | DB 10;    | Length 4174; |  |  |
| Best Local Similarity |     | 100.0%;  | Pred. No. 0;  |           |              |  |  |
| Matches 4174;         |     | Conservative 0;  | Mismatches 0; | Indels 0; | Gaps 0;      |  |  |
| QY                    | 1   | GGCAGAGGGCTGCTCCGATGCTCCAGAGCGGCATGGGCGCCCGCACTGGTGGGACCA          | 60            |           |              |  |  |
| DB                    | 1   | GGCAGAGGGCTGCTCCGATGCTCCAGAGCGGCATGGGCGCCCGCACTGGTGGGACCA          | 60            |           |              |  |  |
| QY                    | 61  | CTGTGGGGCTGCGAGTTCGGAGGTGGATTGGTGGAGGACAACTACACTATCGTGGCTGC        | 120           |           |              |  |  |
| DB                    | 61  | CTGTGGGGCTGCGAGTTCGGAGGTGGATTGGTGGAGGACAACTACACTATCGTGGCTGC        | 120           |           |              |  |  |
| QY                    | 121 | CATTGCCGAGTTCACACACGATCAGCAACGCTCTGTGTTTTTCAATTTACCTCCCATCG        | 180           |           |              |  |  |
| DB                    | 121 | CATTGCCGAGTTCACACACGATCAGCAACGCTCTGTGTTTTTCAATTTACCTCCCATCG        | 180           |           |              |  |  |
| QY                    | 181 | CATGTGCTTGTTCGCCGAGTACGCAACGCTCTGTGTTTTTCAATTTACCTCCCATCG          | 240           |           |              |  |  |
| DB                    | 181 | CATGTGCTTGTTCGCCGAGTACGCAACGCTCTGTGTTTTTCAATTTACCTCCCATCG          | 240           |           |              |  |  |
| QY                    | 241 | GCTCCTAGTGTAGTGGGGATTTGGATCTGTCTACTTCCATGCAACGCTGAGTTTCTGGG        | 300           |           |              |  |  |
| DB                    | 241 | GCTCCTAGTGTAGTGGGGATTTGGATCTGTCTACTTCCATGCAACGCTGAGTTTCTGGG        | 300           |           |              |  |  |
| QY                    | 301 | TCAGATGCTGATGAATCTGCCATTTCTGTGGGTTCGTGATGTGCTTTGGCCATGTGTT         | 360           |           |              |  |  |
| DB                    | 301 | TCAGATGCTGATGAATCTGCCATTTCTGTGGGTTCGTGATGTGCTTTGGCCATGTGTT         | 360           |           |              |  |  |
| QY                    | 361 | TCCCAGGAGGTATTTACCAAGATCTTTCGGAATGACAGGGGACAGGTTCAGGCGAGTGT        | 420           |           |              |  |  |
| DB                    | 361 | TCCCAGGAGGTATTTACCAAGATCTTTCGGAATGACAGGGGACAGGTTCAGGCGAGTGT        | 420           |           |              |  |  |
| QY                    | 421 | GTGTGCTGTGCTGCAATTTACCAAGATCTTTCGGAATGACAGGGGACAGGTTCAGGCGAGTGT    | 480           |           |              |  |  |
| DB                    | 421 | GTGTGCTGTGCTGCAATTTACCAAGATCTTTCGGAATGACAGGGGACAGGTTCAGGCGAGTGT    | 480           |           |              |  |  |
| QY                    | 481 | TTCCCTGATGATCTGGGACTTCCATGCACTGCGCTGTGTTGTCAGAGCTGGAAGGCTG         | 540           |           |              |  |  |
| DB                    | 481 | TTCCCTGATGATCTGGGACTTCCATGCACTGCGCTGTGTTGTCAGAGCTGGAAGGCTG         | 540           |           |              |  |  |
| QY                    | 541 | TGACAAATGTGCTGTGCAATTTACCAAGATCTTTCGGAATGACAGGGGACAGGTTCAGGCGAGTGT | 600           |           |              |  |  |
| DB                    | 541 | TGACAAATGTGCTGTGCAATTTACCAAGATCTTTCGGAATGACAGGGGACAGGTTCAGGCGAGTGT | 600           |           |              |  |  |
| QY                    | 601 | CTTCTGCTGATCAGGACCAAGCCTTCTGTGAGCTGCTCTCTGCGCTCTGCTGCACTCTG        | 660           |           |              |  |  |
| DB                    | 601 | CTTCTGCTGATCAGGACCAAGCCTTCTGTGAGCTGCTCTCTGCGCTCTGCTGCACTCTG        | 660           |           |              |  |  |
| QY                    | 661 | CCTGCATGTGTGGCATATTTCTCATCTGCTTGTCTTCTGTAACCTGGGCTGTGTGCTT         | 720           |           |              |  |  |
| DB                    | 661 | CCTGCATGTGTGTGGCATATTTCTCATCTGCTTGTCTTCTGTAACCTGGGCTGTGTGCTT       | 720           |           |              |  |  |
| QY                    | 721 | GGCTACTTTGATCTGCTCAGAGATACCTGAGCAAGGTCCAGTCAATCTGGCC               | 780           |           |              |  |  |
| DB                    | 721 | GGCTACTTTGATCTGCTCAGAGATACCTGAGCAAGGTCCAGTCAATCTGGCC               | 780           |           |              |  |  |
| QY                    | 781 | CAGCGAGAAATGGGCTTTTATTTGGTGTCCCTTATGTGTCCTTCTGTGTCCTCCCAAGAA       | 840           |           |              |  |  |
| DB                    | 781 | CAGCGAGAAATGGGCTTTTATTTGGTGTCCCTTATGTGTCCTTCTGTGTCCTCCCAAGAA       | 840           |           |              |  |  |
| QY                    | 841 | GTCCGCAAGTCAAGATCAGTGAAGGAGGAGTCAACAGCTTCTCTACTTACTTCTATT          | 900           |           |              |  |  |
| DB                    | 841 | GTCCGCAAGTCAAGATCAGTGAAGGAGGAGTCAACAGCTTCTCTACTTACTTCTATT          | 900           |           |              |  |  |
| QY                    | 901 | CGAGTGGCGCTGGCTTGTGCTAGCAAGATGCTGAGGGGTGAGGAATTTGGTG               | 960           |           |              |  |  |
| DB                    | 901 | CGAGTGGCGCTGGCTTGTGCTAGCAAGATGCTGAGGGGTGAGGAATTTGGTG               | 960           |           |              |  |  |

|    |      |  |      |  |  |  |  |
|----|------|--|------|--|--|--|--|
| QY | 961  | TGCTGTGGGTGTTTAAAAATTCCTCTCTTTGTGATCTAACTGACCACTGTGCTGACCT     | 1020 |  |  |  |  |
| DB | 961  | TGCTGTGGGTGTTTAAAAATTCCTCTCTTTGTGATCTAACTGACCACTGTGCTGACCT     | 1020 |  |  |  |  |
| QY | 1021 | CCCTAGGTTTAACTAGAGGGCTCAGGGACATAAGAGTGTCTTCTCAGTATCTTTCCAGGC   | 1080 |  |  |  |  |
| DB | 1021 | CCCTAGGTTTAACTAGAGGGCTCAGGGACATAAGAGTGTCTTCTCAGTATCTTTCCAGGC   | 1080 |  |  |  |  |
| QY | 1081 | ATACCGGGCTTGTGGGTATGTCCATAATGACATCAACAGAGTAGTCTTTTGGGAGACCT    | 1140 |  |  |  |  |
| DB | 1081 | ATACCGGGCTTGTGGGTATGTCCATAATGACATCAACAGAGTAGTCTTTTGGGAGACCT    | 1140 |  |  |  |  |
| QY | 1141 | AGGGCAACCAAAAGTTTCTGTGAGAGGGTAGTGTCTGTTTCCACCAATTTCCAGACT      | 1200 |  |  |  |  |
| DB | 1141 | AGGGCAACCAAAAGTTTCTGTGAGAGGGTAGTGTCTGTTTCCACCAATTTCCAGACT      | 1200 |  |  |  |  |
| QY | 1201 | CTCAACCCCTCATAAACAGACCTTCTGATTTGGTGATGAAGGTTCCAGAACTTTTCA      | 1260 |  |  |  |  |
| DB | 1201 | CTCAACCCCTCATAAACAGACCTTCTGATTTGGTGATGAAGGTTCCAGAACTTTTCA      | 1260 |  |  |  |  |
| QY | 1261 | TTTTGCCGGGAGAAACTGTCTCTTCAACAAAAACAAAGTGGGCAAAAACACGTGTGGGGGTG | 1320 |  |  |  |  |
| DB | 1261 | TTTTGCCGGGAGAAACTGTCTCTTCAACAAAAACAAAGTGGGCAAAAACACGTGTGGGGGTG | 1320 |  |  |  |  |
| QY | 1321 | TGACTAAGACGGGTGGCTTGTCTGCACTCTTCTAGCTTTTCCAGGTTTCTCCAGTGT      | 1380 |  |  |  |  |
| DB | 1321 | TGACTAAGACGGGTGGCTTGTCTGCACTCTTCTAGCTTTTCCAGGTTTCTCCAGTGT      | 1380 |  |  |  |  |
| QY | 1381 | TGTGATCCTTCCACTACTCTGCTGAGAGTGGAGCCACGGCTTCAGAGGCAAGCTG        | 1440 |  |  |  |  |
| DB | 1381 | TGTGATCCTTCCACTACTCTGCTGAGAGTGGAGCCACGGCTTCAGAGGCAAGCTG        | 1440 |  |  |  |  |
| QY | 1441 | GCAACACCTCTATGCCAAAAGCTACACTCTCTTTAAAGCACACATTTACATAGACACTAT   | 1500 |  |  |  |  |
| DB | 1441 | GCAACACCTCTATGCCAAAAGCTACACTCTCTTTAAAGCACACATTTACATAGACACTAT   | 1500 |  |  |  |  |
| QY | 1501 | TTCTGCTCTTCCAGAGTGCAGAGCTCAGACCCACAGAGAACTCCTTCCAGGTTATGTGAA   | 1560 |  |  |  |  |
| DB | 1501 | TTCTGCTCTTCCAGAGTGCAGAGCTCAGACCCACAGAGAACTCCTTCCAGGTTATGTGAA   | 1560 |  |  |  |  |
| QY | 1561 | GATTCACACACAGCCCCCTCTGTGACCTCTGTGAGAGCAAGGCTCGTTGTAATCGG       | 1620 |  |  |  |  |
| DB | 1561 | GATTCACACACAGCCCCCTCTGTGACCTCTGTGAGAGCAAGGCTCGTTGTAATCGG       | 1620 |  |  |  |  |
| QY | 1621 | GAGGAGCTCTGGAGGCTTGACTGTGAGCGCTGTGAGAGCTTTTGGAAAAGCTTTCTTA     | 1680 |  |  |  |  |
| DB | 1621 | GAGGAGCTCTGGAGGCTTGACTGTGAGCGCTGTGAGAGCTTTTGGAAAAGCTTTCTTA     | 1680 |  |  |  |  |
| QY | 1681 | TTGGAAGCTTTTGGTGTCTTACTCTTCAAGATTTCTGACCCCGTTTCTGCTGTTTTT      | 1740 |  |  |  |  |
| DB | 1681 | TTGGAAGCTTTTGGTGTCTTACTCTTCAAGATTTCTGACCCCGTTTCTGCTGTTTTT      | 1740 |  |  |  |  |
| QY | 1741 | AGGGGATATGCTTCTGAGTGCAGTACCTCTGAGTGTGGGCTTCGGGCTTTTGT          | 1800 |  |  |  |  |
| DB | 1741 | AGGGGATATGCTTCTGAGTGCAGTACCTCTGAGTGTGGGCTTCGGGCTTTTGT          | 1800 |  |  |  |  |
| QY | 1801 | TTTCATAATACCTGGTGTCTTACTGTTTCTCAGAGCACTTTGCTCTTCTGATGCTGGC     | 1860 |  |  |  |  |
| DB | 1801 | TTTCATAATACCTGGTGTCTTACTGTTTCTCAGAGCACTTTGCTCTTCTGATGCTGGC     | 1860 |  |  |  |  |
| QY | 1861 | GGTCACTACACTCTGATTTGGCTGGGATGCCAGGAGGAGGAGGAGGAGCTCTGACA       | 1920 |  |  |  |  |
| DB | 1861 | GGTCACTACACTCTGATTTGGCTGGGATGCCAGGAGGAGGAGGAGGAGCTCTGACA       | 1920 |  |  |  |  |
| QY | 1921 | TGAATGTCTCTCACTGCATCTACTGTCTTCTCACTGGGCTCTCTTCAAGGATCAAGTGG    | 1980 |  |  |  |  |
| DB | 1921 | TGAATGTCTCTCACTGCATCTACTGTCTTCTCACTGGGCTCTCTTCAAGGATCAAGTGG    | 1980 |  |  |  |  |
| QY | 1981 | GTATGGGGATCTCAGGGATGTTTGTAACTCAGGACCTTCTGCTTTCTGACATTTCCATT    | 2040 |  |  |  |  |
| DB | 1981 | GTATGGGGATCTCAGGGATGTTTGTAACTCAGGACCTTCTGCTTTCTGACATTTCCATT    | 2040 |  |  |  |  |



BC059819 4078 bp mRNA linear ROD 21-OCT-2003  
LOCUS Mus musculus cancer related gene-liver 1, mRNA (cDNA clone  
DEFINITION MGC:69583 IMAGE:6839525), complete cds.  
ACCESSION BC059819  
VERSION BC059819.1 GI:37590519  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 4078)  
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Dickson, M.C., Rodriguez, R.W., Touchman, J.W., Green, E.D.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalusz, D.E.,  
Schnitzler, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
12477932  
2 (bases 1 to 4078)  
AUTHORS Strausberg, R.  
DIRECT SUBMISSION  
SUBMITTED (07-OCT-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIB-MGC Project URL: <http://mgc.nci.nih.gov>  
CONTACT: MGC help desk  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library Preparation: M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Neas, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saedi, Jacqueline  
Schein, Duane Smalusz, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranada Tsai, Nataesja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAX Plate: 132 Row: 0 Column: 4  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 21314857.  
Location/Qualifiers  
1. 4078  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="MGC:69583 IMAGE:6839525"  
/tissue\_type="Brain"

gene  
CDS  
/clone\_lib="NIH\_BMAP\_GH0"  
/lab\_host="DHI08"  
/note="Vector: pYX-ASC"  
1. 4078  
/gene="CRG-L1"  
/db\_xref="LocusID:230379"  
76.765  
/gene="CRG-L1"  
/codon\_start=1  
/product="CRG-L1 protein"  
/protein\_id="AAH59819.1"  
/db\_xref="GI:37590520"  
/translation="MGAPHHMDHURAGSSEVDMCEDNNTIVPAIAEFNTISNVLPFI  
LAPICMLFRQYATCFNSGIYLIWLLVVGIGSVYFHALSFLGMDLAILWLWM  
CLAMWFFRRLPKIFRNDRCNDRVFKLGLFSLGWLMTLALFWSIDQAFELSSFH  
PFLYUWVHILICLASVLCVCFAYFDAASBIPQGVIREWSEKNAFIVGPVYVSL  
CAHKKSFKVIT"

ORIGIN

Query Match 92.6%; Score 3864.4; DB 10; Length 4078;  
Best Local Similarity 96.6%; Pred. No. 0;  
Matches 4025; Conservative 0; Mismatches 1; Indels 140; Gaps 2;

QY 10 GCTGCTCGATGCCAGAGCGGCATCGGGCGCCCGCACCTGGTGGACCACTGCGGCG 69  
DB 51 GCTGCTCGATGCCAGAGCGGCATCGGGCGCCCGCACCTGGTGGACCACTGCGGCG 110  
QY 70 TGGCAGTTTCGGAGGTGGATTGGTGGAGGACAACTACACTATCGTGCCTGCCATTGCCGA 129  
DB 111 TGGCAGTTTCGGAGGTGGATTGGTGGAGGACAACTACACTATCGTGCCTGCCATTGCCGA 170  
QY 130 GTTCTACACAGCATCAGCAACGCTCTGTTTTTTCATTTTACCTCCCATCTGCATGTGCTT 189  
DB 171 GTTCTACACAGCATCAGCAACGCTCTGTTTTTTCATTTTACCTCCCATCTGCATGTGCTT 230  
QY 190 GTTCCGCGAGTACCAAGCTGCTTCAACAGCGGCATCTACTTAATATGAGCGCTCCTAGT 249  
DB 231 GTTCCGCGAGTACCAAGCTGCTTCAACAGCGGCATCTACTTAATATGAGCGCTCCTAGT 290  
QY 250 TGTAGTGGGATTGGATCTGCTACTCTCCATCGAACGCTGAGTTTCTCTGGGTTCAGATGCT 309  
DB 291 TGTAGTGGGATTGGATCTGCTACTCTCCATCGAACGCTGAGTTTCTCTGGGTTCAGATGCT 350  
QY 310 TGATGAATCTGCCATCTGTGGGTTCTGATGTGTGCTTTGGCCATGTGGTTTCCAGAG 369  
DB 351 TGATGAATCTGCCATCTGTGGGTTCTGATGTGTGCTTTGGCCATGTGGTTTCCAGAG 410  
QY 370 GTATTTACCAAGATCTTTCGGAATGACAGGGCAGGTTCAAGGCAGTGGTGTGTGCTCT 429  
DB 411 GTATTTACCAAGATCTTTCGGAATGAC-----AGGTGTGCAATGT----- 438  
QY 430 GTCTGCAATTACAACTGCTTTGGCGTTTATCAAGCCCGCCATCAACAATATTTCCCTGAT 489  
DB 439 ----- 438  
QY 490 GATTCTGGACTTCCATGCACTGCGCTGCTGTTGTCAGAGCTGAAGAGTGTGACAATGT 549  
DB 439 -----AGGTGTGCAATGT----- 452  
QY 550 GCGTGTGTTTAACTGGGCTCTTCTCTGGGCTCTGGTGGACTCTGGGCTCTCTTCTGCTG 609  
DB 453 GCGTGTGTTTAACTGGGCTCTTCTCTGGGCTCTGGTGGACTCTGGGCTCTCTTCTGCTG 512  
QY 610 GATCAGGACCAAGCTTCTGTGAGTGTCTCTCTCTCTTTCATCTCCCTACCTGCACTG 669  
DB 513 GATCAGGACCAAGCTTCTGTGAGTGTCTCTCTCTCTTTCATCTCCCTACCTGCACTG 572  
QY 670 TGTGTGCATATTTCTCATCTGCCCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 729  
DB 573 TGTGTGCATATTTCTCATCTGCCCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 632  
QY 730 TGATGCTGCTCAGAGATACCTGAGCAAGGTCCTCAGTTCATCAGATTCTGCCCCCAGGAGAA 789





|    |      |  |      |
|----|------|--|------|
| Qy | 2948 | CTGAGTGCCTCTCTTACGCGGATGGTGTCTCTCAAGTCAAGGAAAGAACCAATTTCTTTT     | 3007 |
| Db | 2953 | CTGAGTGCCTCTCTTACGCGGATGGTGTCTCTCAAGTCAAGGAAAGAACCAATTTCTTTT     | 2912 |
| Qy | 3008 | TTATCAGAAAAGGACTTATTTCCCTTAGGGCCCTCTGCTGACATCCCTAGGAACAGAGATAA   | 3067 |
| Db | 2913 | TTATCAGAAAAGGACTTATTTCCCTTAGGGCCCTCTGCTGACATCCCTAGGAACAGAGATAA   | 2972 |
| Qy | 3068 | AATACGGATGATGTGAATGAACATGCTTGGGTGAAGAAGCCGAGTACCTGACTGAGCC       | 3127 |
| Db | 2973 | AATACGGATGATGTGAATGAACATGCTTGGGTGAAGAAGCCGAGTACCTGACTGAGCC       | 3032 |
| Qy | 3128 | CAGTGGGCCACTTCCCAAGCGAAGCCGGTAAACCGATGTGCATCTCAGAACTTCTCTCCC     | 3187 |
| Db | 3033 | CAGTGGGCCACTTCCCAAGCGAAGCCGGTAAACCGATGTGCATCTCAGAACTTCTCTCCC     | 3092 |
| Qy | 3188 | TTCTGGTAGAGATGACATCGATGAGTGTCTGTGACCACTGGGCCCTGTCATGGGTGTGCG     | 3247 |
| Db | 3093 | TTCTGGTAGAGATGACATCGATGAGTGTCTGTGACCACTGGGCCCTGTCATGGGTGTGCG     | 3152 |
| Qy | 3248 | GGCTCTGTTCTGTTTCCGAATCTACTCAGATCTCAGGACAGAGAAAGCCATGAATGTTAC     | 3307 |
| Db | 3153 | GGCTCTGTTCTGTTTCCGAATCTACTCAGATCTCAGGACAGAGAAAGCCATGAATGTTAC     | 3212 |
| Qy | 3308 | CAAGTGTCTATGGCTGTCTCAGTGATTTTACAGTTTTTGAACCATTAATTGGTTTTTAGGAGAA | 3367 |
| Db | 3213 | CAAGTGTCTATGGCTGTCTCAGTGATTTTACAGTTTTTGAACCATTAATTGGTTTTTAGGAGAA | 3272 |
| Qy | 3368 | TTCTTCTCTCTCTAGTGCCTGTGATGCAAAAGCCAGCCCTTCAAGAAAGTGTCTCTCTCTGT   | 3427 |
| Db | 3273 | TTCTTCTCTCTCTAGTGCCTGTGATGCAAAAGCCAGCCCTTCAAGAAAGTGTCTCTCTCTGT   | 3332 |
| Qy | 3428 | CTGCTTCCCTTTGTAATGTGATCACTCGGGGAAATGGGTGTCTACCCAGGAAAAAGTGC      | 3487 |
| Db | 3333 | CTGCTTCCCTTTGTAATGTGATCACTCGGGGAAATGGGTGTCTACCCAGGAAAAAGTGC      | 3392 |
| Qy | 3488 | CTACCCAGATCACGGGTCAACAGTGTTGTCTGTGAGGATCAAAATGCGCTCTGTGAAAAATA   | 3547 |
| Db | 3393 | CTACCCAGATCACGGGTCAACAGTGTTGTCTGTGAGGATCAAAATGCGCTCTGTGAAAAATA   | 3452 |
| Qy | 3548 | CTGCCACCCATTCTATTTGGTTGGATTTCTCAGGAGTCTGAATCTTCCTCACCAGTCCCTC    | 3607 |
| Db | 3453 | CTGCCACCCATTCTATTTGGTTGGATTTCTCAGGAGTCTGAATCTTCCTCACCAGTCCCTC    | 3512 |
| Qy | 3608 | TTCTCCCAAACCCCTACCCAGAGCAAACACTGGGAATTGAACCTTCTCTGACTCTCTCTTT    | 3667 |
| Db | 3513 | TTCTCCCAAACCCCTACCCAGAGCAAACACTGGGAATTGAACCTTCTCTGACTCTCTCTTT    | 3572 |
| Qy | 3668 | CCCTCAGGTCTGACAACTTAATGGTCTCTGGGGACACCCAGCTAGGGCCTTCCCCAACTCC    | 3727 |
| Db | 3573 | CCCTCAGGTCTGACAACTTAATGGTCTCTGGGGACACCCAGCTAGGGCCTTCCCCAACTCC    | 3632 |
| Qy | 3728 | TTATCCAGCTGAACTTGGATTCTTCCCAAACAGGGCTTGACCTGGGTGTGTTGGTCCCA      | 3787 |
| Db | 3633 | TTATCCAGCTGAACTTGGATTCTTCCCAAACAGGGCTTGACCTGGGTGTGTTGGTCCCA      | 3692 |
| Qy | 3788 | CTGGCCAAACACATCTTTTGGCCAGATTGGGAATCTCAATAGATTTTATAGACATTAAT      | 3847 |
| Db | 3693 | CTGGCCAAACACATCTTTTGGCCAGATTGGGAATCTCAATAGATTTTATAGACATTAAT      | 3752 |
| Qy | 3848 | CTCCCAAGACTTTAAACATGGCTGTGCTTTTCCATACACATCCGGTCAGATTTTAAAAA      | 3907 |
| Db | 3753 | CTCCCAAGACTTTAAACATGGCTGTGCTTTTCCATACACATCCGGTCAGATTTTAAAAA      | 3812 |
| Qy | 3908 | CTATTTTATAACCAAGGAATTAACCAAGCAAAATAGAGTACTTTTCAGATATAAACTGTG     | 3967 |
| Db | 3813 | CTATTTTATAACCAAGGAATTAACCAAGCAAAATAGAGTACTTTTCAGATATAAACTGTG     | 3872 |
| Qy | 3968 | TTTTCATCTTTATGTAGAGTGTGCTATGTATATAGCGGTATGTACCTGTGCTGAAGTAATA    | 4027 |
| Db | 3873 | TTTTCATCTTTATGTAGAGTGTGCTATGTATATAGCGGTATGTACCTGTGCTGAAGTAATA    | 3932 |

|    |      |   |      |
|----|------|---|------|
| QY | 4028 | TTAACCATAGCTCTGGGAGGATTTACAGACCTTTTGGACCTTTTAAGCTTTTGTGGAATC    | 4087 |
| Db | 3933 | TTAACCATAGCTCTGGGAGGATTTACAGACCTTTTGGACCTTTTAAGCTTTTGTGGAATC    | 3992 |
| QY | 4088 | TGATAACCATGGTCAATATTAAAGCCCAATAACTGGGCAATTTCTGTGAATAAACAATGCATA | 4147 |
| Db | 3993 | TGATAACCATGGTCAATATTAAAGCCCAATAACTGGGCAATTTCTGTGAATAAACAATGCATA | 4052 |
| QY | 4148 | TGTATCTAAAAAAAAAAAAAAAAAAAA                                     | 4173 |
| Db | 4053 | TGTATCTAAAAAAAAAAAAAAAAAAAA                                     | 4078 |

RESULT 3  
BX005084

```

LOCUS       EX005084               90370 bp    DNA    linear   ROD 23-OCT-2003
DEFINITION  Mouse DNA sequence from clone RP24-468M3 on chromosome 4, complete
            sequence.
ACCESSION   EX005084
VERSION     EX005084.16  GI:37936459
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
            Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 90370)
AUTHORS    Barlow,K.
TITLE      Direct Submission
JOURNAL     Submitted (22-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquerry@sanger.ac.uk
            On Oct 23, 2003 this sequence version replaced gi:37693442.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence. Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            ----- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquerry@sanger.ac.uk

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e. paired quality = 30); an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP24-468M3 is from a Male (C57BL/6J) mouse SAC Library VECTOR: 07AR8AC1.

FEATURES  
SOURCE

```

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP24-468M3"
/clone_lib="RPCI-24"

```

## ORIGIN

Query Match 83.1%; Score 3467.6; DB 10; Length 90370;

| Best Local Similarity 99.5%; Pred. No. 0;                      |       |  |       |
|--|-------|--|-------|
| Matches 3488; Conservative 0; Mismatches 14; Indels 2; Gaps 1; |       |  |       |
| QY   | 654   | TCCCTCACTGACACGTGTGGGCAATATCTCATCTGCGCTTCTGCTCGTACCTGGGCTGTG   | 713   |
| Db   | 78025 | TCACATCTCCCTCTCTTTGGAGGCAATATCTCATCTGCGCTTCTGCTCGTACCTGGGCTGTG | 78084 |
| QY   | 714   | TGTGCTTCGCGCTACTTTGATGCTGCGCTCAGAGATACCTGAGCAAGGTCAGATCAGAT    | 773   |
| Db   | 78085 | IGTGCTTCGCGCTACTTTGATGCTGCGCTCAGAGATACCTGAGCAAGGTCAGATCAGAT    | 78144 |
| QY   | 774   | TCTGCCCAAGGAGAAATGGGCTTTTATTTGGTGTCCCTTATGTGTCCCTTCTGTGTGCC    | 833   |
| Db   | 78145 | TCTGCCCAAGGAGAAATGGGCTTTTATTTGGTGTCCCTTATGTGTCCCTTCTGTGTGCC    | 78204 |
| QY   | 834   | ACAAGAAGTGCACGTCGAAGATCAGTGTGCAAGGAGGAGTGCACAGCTTCTCTACTTAC    | 893   |
| Db   | 78205 | ACAAGAAGTGCACGTCGAAGATCAGTGTGCAAGGAGGAGTGCACAGCTTCTCTACTTAC    | 78264 |
| QY   | 894   | TTCTATTGAGTGCCTGCGCTGGGCTTCTGTTGCTAGCAAGATGGCTGAGGGGTTGAGGA    | 953   |
| Db   | 78265 | TTCTATTGAGTGCCTGCGCTGGGCTTCTGTTGCTAGCAAGATGGCTGAGGGGTTGAGGA    | 78324 |
| QY   | 954   | ATTGCTGTGCTGGGTGTTTAAATTTCTGCTCTCTTTGTGATCTAACTGGACCACTGTGC    | 1013  |
| Db   | 78325 | ATTGCTGTGCTGGGTGTTTAAATTTCTGCTCTCTTTGTGATCTAACTGGACCACTGTGC    | 78384 |
| QY   | 1014  | CTGACCTCCCTAGGTTTAACTGAGGGCTCAGGGACATAAGGTGTCTTCTCAGTATCCTT    | 1073  |
| Db   | 78385 | CTGACCTCCCTAGGTTTAACTGAGGGCTCAGGGACATAAGGTGTCTTCTCAGTATCCTT    | 78444 |
| QY   | 1074  | TCCAGGCATACGGGCTTGTCTGGGTTATGTGCCAATATGACATCAACAGAGTAGTCTTTGG  | 1133  |
| Db   | 78445 | TCCAGGCATACGGGCTTGTCTGGGTTATGTGCCAATATGACATCAACAGAGTAGTCTTTGG  | 78504 |
| QY   | 1134  | GAGACTAGGGCAACCCAAAGTTTCTGTCTGAGAGGGTAGCTTGTGTTCCACCATTT       | 1193  |
| Db   | 78505 | GAGACTAGGGCAACCCAAAGTTTCTGTCTGAGAGGGTAGCTTGTGTTCCACCATTT       | 78564 |
| QY   | 1194  | CCAGACTCTCAACCCCTCATAAACAAGACCTCTGTATTTGGTGTGATGAAGGTTCCAGAA   | 1253  |
| Db   | 78565 | CCAGACTCTCAACCCCTCATAAACAAGACCTCTGTATTTGGTGTGATGAAGGTTCCAGAA   | 78624 |
| QY   | 1254  | CTTTTCATTTGCGGGAGAACTGTCTTCAACAAAACCAAGTGGGCAAAAACAGGTGTG      | 1313  |
| Db   | 78625 | CTTTTCATTTGCGGGAGAACTGTCTTCAACAAAACCAAGTGGGCAAAAACAGGTGTG      | 78684 |
| QY   | 1314  | GGGGGTGTGACTTAAGACGGGTGTGTCACTCGACCTCTTAGCTTTCCAGGTTCTCC       | 1373  |
| Db   | 78685 | GGGGGTGTGACTTAAGACGGGTGTGTCACTCGACCTCTTAGCTTTCCAGGTTCTCC       | 78744 |
| QY   | 1374  | AGTGTGTTGGATCTTCCACTCTCTGCTGAGAGTGGAGCCACGGCTTCAGAGGC          | 1433  |
| Db   | 78745 | AGTGTGTTGGATCTTCCACTCTCTGCTGAGAGTGGAGCCACGGCTTCAGAGGC          | 78804 |
| QY   | 1434  | AAAGCTGGCAACCCCTCTATGCAAAAGCTACACTCTCTTTAAGCACACATTAAC         | 1491  |
| Db   | 78805 | AAAGCTGGCAACCCCTCTATGCAAAAGCTACACTCTCTTTAAGCACACATTAACAT       | 78864 |
| QY   | 1492  | AGACACTATTTCTGCTTCCAGAGTGCAGAGCCTCAGACCCACAGAGAACTCTTCAGG      | 1551  |
| Db   | 78865 | AGACACTATTTCTGCTTCCAGAGTGCAGAGCCTCAGACCCACAGAGAACTCTTCAGG      | 78924 |
| QY   | 1552  | TTATGTGAGATTCACACACAGCCCCCTTGTGTGACCTCTGTGAGCAAGGCCCTGTT       | 1611  |
| Db   | 78925 | TTATGTGAGATTCACACACAGCCCCCTTGTGTGACCTCTGTGAGCAAGGCCCTGTT       | 78984 |
| QY   | 1612  | GTAATCGGAGGACGCTCTGGAGGCTGTGACTGTGAGCGCCTGTGAAGATTTGGAAG       | 1671  |
| Db   | 78985 | GTAATCGGAGGACGCTCTGGAGGCTGTGACTGTGAGCGCCTGTGAAGATTTGGAAG       | 79044 |
| QY   | 1672  | CTTTTCTTATTGGAAGCTTTTGGTGTGTGCTTACCTCCAGATTTCTGACCCCGTTTCTCT   | 1731  |





QY 2332 TTCAATAGCAATTTACCAACTCTCCAGGTAGCAAAACACAGTCAATTTGATGGGATTTA 2391  
Db 43911 TTCAATAGCAATTTACCAACTCTCCAGGTAGCAAAACACAGTCAATTTGATGGGATTTA 43970  
QY 2392 CAAAGCAGGAATCTTGGTTTGAATGATGTATATGTATATTTACTGTAGCTTAAATA 2451  
Db 43971 CAAAGCAGGAATCTTGGTTTGAATGATGTATATGTATATTTACTGTAGCTTAAATA 44030  
QY 2452 TGTTTAAATGACTTTGAGTGAATCTTTGTGGCAACCTTAGGGTTTATGGATCGAAT 2511  
Db 44031 TGTTTAAATGACTTTGAGTGAATCTTTGTGGCAACCTTAGGGTTTATGGATCGAAT 44090  
QY 2512 GCGGCTGAGTCTCTTTGGTTTGGTCTAGAGTTTTCAGAGGGCAAAATCAAACTAGCC 2571  
Db 44091 GCGGCTGAGTCTCTTTGGTTTGGTCTAGAGTTTTCAGAGGGCAAAATCAAACTAGCC 44150  
QY 2572 TTGTTTCTGGTTCATCTTACAGGCTCGGAGCTGCCCTTATATTTTACATAAGGACT 2631  
Db 44151 TTGTTTCTGGTTCATCTTACAGGCTCGGAGCTGCCCTTATATTTTACATAAGGACT 44210  
QY 2632 TATTATACATAAGTCTGTATAAATGCTTGAAGATGACACCTTAGCTTCTCTCCAG 2691  
Db 44211 TATTATACATAAGTCTGTATAAATGCTTGAAGATGACACCTTAGCTTCTCTCCAG 44270  
QY 2692 GGGTCGTCTGGGGCTGGAGTTGGTTCAACAGTTTAAAGAACCCGTGTTGCTCTCTCCAG 2751  
Db 44271 GGGTCGTCTGGGGCTGGAGTTGGTTCAACAGTTTAAAGAACCCGTGTTGCTCTCTCCAG 44330  
QY 2752 AGCAACCCAGCGACTTGTGAGGCTCATACAGCAACTCAGCCCTGGGCACTCAGCGC 2811  
Db 44331 AGCAACCCAGCGACTTGTGAGGCTCATACAGCAACTCAGCCCTGGGCACTCAGCGC 44390  
QY 2812 CTCCTCTGACTTCTAAGGCACTTGGTGACAGCTTACATACAGGCAAGCAAACTTAT 2871  
Db 44391 CTCCTCTGACTTCTAAGGCACTTGGTGACAGCTTACATACAGGCAAGCAAACTTAT 44450  
QY 2872 ACATGTAAACGTAATAAATGCTAAGTGTAGTACAGCGCTGAAGAAAGAGTTTTAGATG 2931  
Db 44451 ACATGTAAACGTAATAAATGCTAAGTGTAGTACAGCGCTGAAGAAAGAGTTTTAGATG 44510  
QY 2932 CAAAGTTTAGTCTGACCTGAGTGCCTTCTTACCGCGATGGTGTCTCTCAAGTCAAGG 2991  
Db 44511 CAAAGTTTAGTCTGACCTGAGTGCCTTCTTACCGCGATGGTGTCTCTCAAGTCAAGG 44570  
QY 2992 AAGAACCAATTTCTTTTATCAGGAABAGGACTTATTCCTTAGGCGCTCTGCTGACATCC 3051  
Db 44571 AAGAACCAATTTCTTTTATCAGGAABAGGACTTATTCCTTAGGCGCTCTGCTGACATCC 44630  
QY 3052 CTAGGAACAGAGATAAAATACGGATGGATGTGAATGAACATGCTTGGGTGAAGAGCGCG 3111  
Db 44631 CTAGGAACAGAGATAAAATACGGATGGATGTGAATGAACATGCTTGGGTGAAGAGCGCG 44690  
QY 3112 AGTACCTGACCTGGACCCAGTGGGCCACTTCCACAGCGAAGCCCGTAAACGATGTGAC 3171  
Db 44691 AGTACCTGACCTGGACCCAGTGGGCCACTTCCACAGCGAAGCCCGTAAACGATGTGAC 44750  
QY 3172 TCCAGAACTTCTCTCCCTTCTCGTAGAGATGACATCGATGATGTGTCGTCGACCTGGG 3231  
Db 44751 TCCAGAACTTCTCTCCCTTCTCGTAGAGATGACATCGATGATGTGTCGTCGACCTGGG 44810  
QY 3232 CCGTGCATGGGTGCGGCTCTGTTCTGTTCCGAACTCTACCTGAGATCTCAGGACAGAG 3291  
Db 44811 CCGTGCATGGGTGCGGCTCTGTTCTGTTCCGAACTCTACCTGAGATCTCAGGACAGAG 44870  
QY 3292 AAGCCATGAATGTTACCAAGTGGTCATGGCTGTGAGTATTTACAGTTTGAACCAATTA 3351  
Db 44871 AAGCCATGAATGTTACCAAGTGGTCATGGCTGTGAGTATTTACAGTTTGAACCAATTA 44930  
QY 3352 TTGGTTTTTAGGAAATTTCTTCTCTCTAGTGCCTGTGATGCAAGGCAAGCCCTTCA 3411  
Db 44931 TTGGTTTTTAGGAAATTTCTTCTCTCTCTAGTGCCTGTGATGCAAGGCAAGCCCTTCA 44990

QY 3412 GAAGTGTTCCTTCTGTCTGCTTCCCTTTGTAATGTGATCCACTCGGGGAAATGGGTGTCT 3471  
Db 44991 GAAGTGTTCCTTCTGTCTGCTTCCCTTTGTAATGTGATCCACTCGGGGAAATGGGTGTCT 45050  
QY 3472 ACCCAGGGAAGAGTGCCTACCCAGTACAGGGTCAACAGTGTGTGTTGAGGATCAACA 3531  
Db 45051 ACCCAGGGAAGAGTGCCTACCCAGTACAGGGTCAACAGTGTGTGTTGAGGATCAACA 45110  
QY 3532 TGGCTCTGTGAAATACTACCCACCTTCTATTTGGTTGGATTTCTCAGAGTCTGAATCT 3591  
Db 45111 TGGCTCTGTGAAATACTACCCACCTTCTATTTGGTTGGATTTCTCAGAGTCTGAATCT 45170  
QY 3592 TCCCTCAGAGTCTCTTCTTCCCAACCCCTACCCAGAGCAACACTCGGGATTTGAACCTT 3651  
Db 45171 TCCCTCAGAGTCTCTTCTTCCCAACCCCTACCCAGAGCAACACTCGGGATTTGAACCTT 45230  
QY 3652 CTCTGACTCTCTTCTTCCCTCAGGTCTGACAACTAATGTCTCTGGGACACCCAGCTAG 3711  
Db 45231 CTCTGACTCTCTTCTTCCCTCAGGTCTGACAACTAATGTCTCTGGGACACCCAGCTAG 45290  
QY 3712 GGCCTTCCCAACTCTTATCCAGCTGAACCTTGGATTTCTTCCCAACAGGGCTTGAACCTG 3771  
Db 45291 GGCCTTCCCAACTCTTATCCAGCTGAACCTTGGATTTCTTCCCAACAGGGCTTGAACCTG 45350  
QY 3772 GGTGCTGTGGTCCCACTGGCCCAACACACATCTTTGGCCAGATTGGGATTTCTCAATAGA 3831  
Db 45351 GGTGCTGTGGTCCCACTGGCCCAACACACATCTTTGGCCAGATTGGGATTTCTCAATAGA 45410  
QY 3832 TTTTATAGACATTTATCTTCCACAGACTTTTAAACATGCTTGTCTTCCATACACAT 3891  
Db 45411 TTTTATAGACATTTATCTTCCACAGACTTTTAAACATGCTTGTCTTCCATACACAT 45470  
QY 3892 CCGGTCAAGTTTAAACATTTTATTAACCAACAGAGAAATTAACCAAGCAATAGAGTACTT 3951  
Db 45471 CCGGTCAAGTTTAAACATTTTATTAACCAACAGAGAAATTAACCAAGCAATAGAGTACTT 45530  
QY 3952 TCAGATATAAATCTGTTTTCATCTTATGTAGAGTGTCTATGTATAGGCGGTATGTAC 4011  
Db 45531 TCAGATATAAATCTGTTTTCATCTTATGTAGAGTGTCTATGTATAGGCGGTATGTAC 45590  
QY 4012 CCGTCTGAAGTAAATTAATTAACCATAGCTCTCGGAGGATTTACAGACCTTTTGCACCTTAT 4071  
Db 45591 CCGTCTGAAGTAAATTAATTAACCATAGCTCTCGGAGGATTTACAGACCTTTTGCACCTTAT 45650  
QY 4072 GCTTTTGTGAATCTGTGAATACCATGTCAATATTAAGCCAAATACTGGCATTTTCTG 4131  
Db 45651 GCTTTTGTGAATCTGTGAATACCATGTCAATATTAAGCCAAATACTGGCATTTTCTG 45710  
QY 4132 TGAATAACATGCATATGTATCTA 4155  
Db 45711 TGAATAACATGCATATGTATCTA 45734

## RESULT 5

AC120245

LOCUS

DEFINITION

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

AC120245

Rattus norvegicus clone CH230-259012, WORKING DRAFT SEQUENCE, 5

unordered pieces.

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

184163 bp DNA linear HTG 15-NOV-2002

Rattus norvegicus clone CH230-259012, WORKING DRAFT SEQUENCE, 5

unordered pieces.

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

184163 bp DNA linear HTG 15-NOV-2002

Rattus norvegicus clone CH230-259012, WORKING DRAFT SEQUENCE, 5

unordered pieces.

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

184163 bp DNA linear HTG 15-NOV-2002

Rattus norvegicus clone CH230-259012, WORKING DRAFT SEQUENCE, 5

unordered pieces.

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

184163 bp DNA linear HTG 15-NOV-2002

Rattus norvegicus clone CH230-259012, WORKING DRAFT SEQUENCE, 5

unordered pieces.

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

184163 bp DNA linear HTG 15-NOV-2002

Rattus norvegicus clone CH230-259012, WORKING DRAFT SEQUENCE, 5

unordered pieces.

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

184163 bp DNA linear HTG 15-NOV-2002

Rattus norvegicus clone CH230-259012, WORKING DRAFT SEQUENCE, 5

unordered pieces.

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

184163 bp DNA linear HTG 15-NOV-2002

Rattus norvegicus clone CH230-259012, WORKING DRAFT SEQUENCE, 5

unordered pieces.

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

184163 bp DNA linear HTG 15-NOV-2002

Rattus norvegicus clone CH230-259012, WORKING DRAFT SEQUENCE, 5

unordered pieces.

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

184163 bp DNA linear HTG 15-NOV-2002

Rattus norvegicus clone CH230-259012, WORKING DRAFT SEQUENCE, 5

unordered pieces.

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

184163 bp DNA linear HTG 15-NOV-2002

Rattus norvegicus clone CH230-259012, WORKING DRAFT SEQUENCE, 5

unordered pieces.

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

Rattus norvegicus

184163 bp DNA linear HTG 15-NOV-2002

Rattus norvegicus clone CH230-259012, WORKING DRAFT SEQUENCE, 5&lt;/

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensheva, L., Loulseged, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwackeleme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plummer, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wood, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Unpublished

Worley, K.C.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (05-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 184163)  
Rat Genome Sequencing Consortium.  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 15, 2002 this sequence version replaced gi:22856351.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GVT

Center clone name: CH230-259012

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 175276 bases at least Q40

Consensus quality: 176285 bases at least Q30

Consensus quality: 177113 bases at least Q20

Estimated insert size: 175972; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 5 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 177885: contig of 177885 bp in length

\* 177886 177885: gap of unknown length

\* 177986 179377: contig of 1392 bp in length

\* 179378 179477: gap of unknown length

\* 179478 180571: contig of 1094 bp in length

\* 180572 180571: gap of unknown length

\* 180672 182336: contig of 1565 bp in length

\* 182337 182336: gap of unknown length

\* 182337 184163: contig of 1827 bp in length.

----- Location/Qualifiers

1..184163

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-259012"

1..1513

/note="wgs end extension"

clone\_end:Sp6"

complement(4969..5842)

/note="clone boundary"

clone\_end:Sp6"

site:

end\_sequence:B2169793"

40368..40445

/note="clone boundary"

clone\_end:T7"

site:

end\_sequence:B2169790"

171006..172144

/note="wgs end extension"

clone\_end:T7"

173214..174400

/note="wgs end extension"

clone\_end:T7"

175979..177885

/note="wgs end extension"

clone\_end:T7"

ORIGIN

Query Match

Best Local Similarity

Matches 1886; Conservative

0; Mismatches 351; Indels 203; Gaps 25;

QY 632 GAGCTCTCTCCCTTTCACCTTCCCTACCTGCGCATATTCATCTGC 691

Db 12892 GAGAGCATGCGAGCACCCCGCTCACCCCTCCCTTGCGGACATCTCATCTGC 12951

QY 692 CTTCGCTTCGTACCTGGGCTGTGTGCTTCGGCTACTTTGATGCTCGCTCAGAGATACCT 751

Db 12952 CTTCGCGCGTACTTGGGCTGGCTGCTTCGCTACTTCGATGCTGCTCAGAGATACCC 13011



QY 752 GAGCAAGTCCAGTTCATCAGATTCCTGCCCCAGCAGAAATGGCTTTTATTTGGTGTCCCT 811  
Db 13012 GAACAGGTCCAGTTCATCAGATTCCTGCCCCAGTGAATAATGGCTTTTATCGGTGTCCCC 13071  
QY 812 TATGTGTCCTCTCTGTGCCCCCAAGAAGTCCGCCAGTCAAGATCACGTCGATGGCAAGGC 871  
Db 13072 TATGTGTCCTCTCTGTGCCCCCAAGAAGTCCGCCAGTCAAGATCACGTCGATGGCAAGGC 13131  
QY 872 AGTGACCAAGTTCCT 931  
Db 13132 AGTGACCAAGTTCCT 991  
QY 932 AGATGGCTGAGGGGTTGAGGAATTTGATGTAGTGTGGGTGTTTAAATTTCTCTCTCT 13191  
Db 13192 AGATGGCTGAGGGGTTGAGGAATTTGATGTAGTGTGGGTGTTTAAATTTCTCTCTCT 13251  
QY 992 TGATCTAACTGGACCACTGTGCTGACCTCCCTAGGTAAAGTA---GAGGGCTCAGGGAC 1048  
Db 13252 TGATCTAACTGGACCACTGTGCTGACATCCCAAGGTGGTAAATGGAGGGCTCAGGGAC 13311  
QY 1049 ATAAG-GTGTCT 1107  
Db 13312 ATGAGTGGGTGGGCTCGAGCACTTTTACGGGCACATCGGGCTGTGCTGTGTATATCCATG 13371  
QY 1108 ATGACATCAACAGAGTAGTTCCTTGGGAGACCTAGGGCAACCCAAAGTTTCTTGTCTGGAG 1167  
Db 13372 ATGACATTTGACAGAGTAGTTCCTTGGGAGACCCAGAGAGCCCAAGCTTCTTGTCTGGAA 13431  
QY 1168 AGGTGAGCTTGG---TGTTTCCACATTTCCAGACTCTCAACCCCTGATAAACAAGACC 1224  
Db 13432 AAGTGGCTTGCCATAGCCCCACCATTTCCAGGCTCTCTCATTAACAGAGTCAAGGCC 13491  
QY 1225 TTCTGATTTGGTATGAAGTTCCAGAACTTTTCAATTTGCGGGAGAAACTGTCTCTTC 1284  
Db 13492 C-CTGATTTCTGTATAGAGTTCCAGAACTTTTCAATTTGCTGGGAGAAACTGTCTCTTC 13550  
QY 1285 AACAAACCAAGTGGGCAAA----- 1304  
Db 13551 AACAAATCGAGTAGGCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAA 13610  
QY 1305 -----ACACGTGTGGG 1315  
Db 13611 AACAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAATGG 13670  
QY 1316 GGGTGTGACTAAGACGGGTGGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1375  
Db 13671 GGGTGTGACTAAGACGGGTGGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 13722  
QY 1376 GTGTTGTGGATCCTTCCACTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1435  
Db 13723 ATGCTGTGGATCCTTCCACTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 13782  
QY 1436 AGCTGGCAACACCTCTATGCAAAAGTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1493  
Db 13783 AGCCAGTAACACCTCTACGCCAAGAGTACAGTCCACTTTAAGCACACATTAACAACCTG 13842  
QY 1494 ACATATTTCT 1553  
Db 13843 ACATATGTCT 13900  
QY 1554 ATGTGAAGATTCACACACAGCCCTCT-----CTTGTGACCTCTGTGAGAGCAAGG 1604  
Db 13901 -TGTGGCAATTCCACACATAGCCCCCTACCGCTTGCCACTTGACCTCTGTGAGAGCAAGG 13959  
QY 1605 CTTGTTGTAATCCGGAGGAGCTCTG-GAGGCTTGACTGTGGAGCGCTCTGTAAGATT 1663  
Db 13960 ACTCAITGTCACTCTGAGGAGCTCTGAGGGTTTGTGATGTAGAGCTCTCTGTGAAGCTT 14019  
QY 1664 TTGG-AAAGCTTTCTTATTTGGAAGCTTTTGGTGTGCTGTACCTCCCAAGATTCAGACC 1722  
Db 14020 TTGGAAGCTTTCTTATTTGGAAGCTTACAGTGTGTTTACCTCTCCAGACTCTGACCC 14079

QY 1723 CGTTTCTCTGCTGTTTATAGGGGATATGTGCTTCTCTGAGTGACGTAGCCTCCCTAGGATG 1782  
Db 14080 CATTTTCTCTGCTGTTTTCAGGGGATGTGTGCTTCTCTGAGGATGCCACTTCTCCCTGGATG 14139  
QY 1783 TGGGCTCCGGCTTGTGTTTTCATAATACCTGTGTCTAACTAGTCTTCTCAGAGCACTTTG 1842  
Db 14140 TGGACCTCCAGCTTATTTTGTATATACC--GGTGTAACTAGTCTTCTCAGAGCCCTTTG 14198  
QY 1843 CTCTCTTTGATGTCTGGGCGGTCACTACACTCTGATTTGGCTGGCT--GGGATCCCCAGGAG 1901  
Db 14199 CTCTCTTTGATGTCTGGCCAGTCACGACACTCTGATTTGGCTGGCTAGGGGTCACAGGTAG 14258  
QY 1902 GAAGGGGAGACTCTGACATGAATGTCTCTCACCTGTCATCTCTCTCTCTCTCTCTCTCTCT 1961  
Db 14259 GAAGGGGAGACTCTGACCTGAATGTCTCTCACCTGTCATCTCTCTCTCTCTCTCTCTCTCT 14318  
QY 1962 TCCTTCAGGGTATGAATGGGTATGGG--GTATCTCAGGGATGTTTGTAACTCAGGCACCTT 2020  
Db 14319 TCTCAGGGTATGAGTGGGAATGGGTATATCTCAGGGATGTTTGTAACTCAGGCACCTT 14378  
QY 2021 CTGCTTTCTGACATTTCCATTGTCTAGTGTGTGAGACTCTCAGGGGACTGAGTGTGAAC 2080  
Db 14379 CTGCTTTCTGACATTTCCATTGT--TGGGGTGGGACAGCGCTCAGGGGACTGCAAGTTGCTAC 14437  
QY 2081 ACCGTTCTGAGTGTGCTCTGAGTCTGAAGAGCTAGCTGCGGGTCTCTGCGACTTCTAGG 2140  
Db 14438 ACTGTTCTGAGTGGGCTCTGAGTTTGAAGAGTTGGC---CGGTTCTGGCGCTTCCAGG 14494  
QY 2141 ATCTCTTACTCTGTTTGAACCTTTCACAGGGTCAAAAGTGGGAACTGACCTTAAAGAGTT 2200  
Db 14495 AACTCTTACTCCGTATAGAAGCTTTCACAGGGTCCAAAGTGGGAACTGACT--AGGAGTT 14552  
QY 2201 TTTAAATGAATAGACTTCATTCGCTTTTGTGGCTTTGAGTGTAAAGTAACCTTGTCTG 2260  
Db 14553 TTTAAATGCATAGGCTTCATTCGCTTCTGGGGCTTTGAACTTTTGAAGTAGCTTGTCTA 14612  
QY 2261 GGCTATTTCTGTTTATAGTCAAAATATTTATAGAACATGAAGTGTAAATAGTTGT 2320  
Db 14613 GGCTCTTCTTGTCTGTAGTCAAAATATCTGTAGAACCCGAAATGTGTGAAGAGAGTCAAT 14672  
QY 2321 CTTTTATTAATTCATAGCAATTTACCAACTCTCCAGGTAGCAACACACAGTC---AT 2376  
Db 14673 CTTTTATTAAGTTTATAGCAATTTACTAACTCTCCAGGTAGCAAGCCATAGTCAATTTAT 14732  
QY 2377 TTGATTTGGGATTTACAGGAGGAGAACTACTGTTTGAATGTATGTATATGTCAATTTA 2436  
Db 14733 TTGATTTGGGATTTATAAGGCAGAGATACTTGGCTTTGAATGACTGTATATGTTATTTA 14792  
QY 2437 CTGTAGCTTAAATATCTTTAAATGACTTTGAGTGAATGTTTGTGGCAACCTA--GGG 2494  
Db 14793 TTGCAGCTTAAATATATTTTAAATGATTTTGAATGTTTGTGGCAACCTAGGGGG 14852  
QY 2495 GTTTATGGATCAGAAATTCGCGCTGAGTCTTGTGTTTGTGAGTCTAGAAGTTTTCAGAGG 2554  
Db 14853 GTTCATGGACAGAAATTCGCTGAGTCTTGTGTTTGTGAGTCTAGAATTTTTCAG--- 14909  
QY 2555 GCATAATCAAACTAGCCTTGTCTGTTTCACTTACAGGCTCGGAGCTGCGCTTATA 2614  
Db 14910 -----ACTAGCCTTGGCTTCCAGTCTATCTCAGCAGGCAAGGGGCTGCTCTATA 14959  
QY 2615 TATTCTACATAAGGACTTATTATACATAAGTCTGTATAAATGCTCTCAAGATGACACTA 2674  
Db 14960 TATTCTACA-----CGNATATACATCTCAGGATGACCGCA 14997  
QY 2675 GCTGCTCTCATCTGGAAGGCTC-----GTCTGGGCTGGAGAGTTTGGTTCAACAGTTAA 2728  
Db 14998 GCTGCTCTCTCTGAAAGGCGCGAGCTGTCTGGGCTGGAGAGCTGTTTCAGCCGATAA 15057  
QY 2729 GRACCCGTGTTGCTCTCTCCAG-----ACGACCCAGGACTTGTGAGGCTCAT 2777  
Db 15058 AAACACATGTTGGCTCTCCAGAGACCCAGAGTTTGTATTCGACATCTTATAAAGCTGAC 15117  
QY 2778 AACAGCAATCCAGCCCTGGGGCATCTGACGCCCTCTTCTGACTTCTTAAGG----- 2829

|           |  |
|-----------|--|
| TITLE     | Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., |
| JOURNAL   | Weinstock, G. and Gibbs, R. A.   |
| REFERENCE | Direct Submission  |
| AUTHORS   | Unpublished  |
| TITLE     | 2 (bases 1 to 249734)  |
| JOURNAL   | Worley, K. C.  |
| REFERENCE | Direct Submission  |
| AUTHORS   | Submitted (17-OCT-2001) Human Genome Sequencing Center, Department     |
| TITLE     | of Molecular and Human Genetics, Baylor College of Medicine, One       |
| JOURNAL   | Baylor Plaza, Houston, TX 77030, USA                                   |

AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly. In location scaffold1 within each contig, scaffold

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole

```

genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAXA
Center clone name: CH230-3E15
----- Summary Statistics
Assembly program: Atlas;
Consensus quality: 241576 bases at least Q40
Consensus quality: 243506 bases at least Q30
Consensus quality: 245010 bases at least Q20
Estimated insert size: 253391; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank draft data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 248465: contig of 248465 bp in length
* 248466 248565: gap of unknown length
* 248566 249734: contig of 1169 bp in length.
FEATURES
      Location/Qualifiers
          1..249734
              /organism="Rattus norvegicus"
              /mol_type="genomic DNA"
              /db_xref="taxon:10116"
              /clone="CH230-3E15"
          1..2605
              /note="wgs_contig"
      misc_feature
ORIGIN
Query Match          29.3%; Score 1222.4; DB 2; Length 249734;
Best Local Similarity 77.3%; Pred. NO. 0;
Matches 1886; Conservative 0; Mismatches 351; Indels 203; Gaps 25;

```



```

Db 150993 GCTGCTTTGCTGAAGGCGCCGAGTGTCTGCGGCTGGAGAGCTGGTTTCAGCCGATAA 151052
QY 2729 GAAACCGTGTGCTCTCTCCAG-----ACGACCCAGCGACTTGTGAGGCTCAT 2777
Db 151053 AAGACATGTTGCTCTCCAGGACCCAGGTTGATTTCCAGCATCTTATAAGAGTGAC 151112
QY 2778 AACAGCAACTCCAGCCCTGGGGCATCTGACGCCCTCTTCTGACTTCTAAG----- 2829
Db 151113 AACCTGTAGCTCAGTCCAGGGGATCCAGTGCCTCTTCTGACCTCCAAGGGCCAGAG 151172
QY 2830 ---GCATTGGTGGCACCTACATACAGGCGGCAAAACATTATACATGTAACGTAATA 2886
Db 151173 CATTGATGGTACACCTATGATCATACGAGCGCAAGTCAATTCGTACATGTAAGT---- 151228
QY 2887 AATGCATAAGTTAGTGAGACGGCTGAAGGAAAGGAGTTTT 2926
Db 151229 AATGAGTAAGTTAATAAGCGCGGAGGAAAGGTTTTT 151268

RESULT 7
LOCUS CQ842051 2852 bp DNA linear PAT 02-AUG-2004
DEFINITION Sequence 698 from Patent EP1440981.
ACCESSION CQ842051
VERSION CQ842051.1 GI:50893838
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamamoto,J., Isono,Y., Nagai,K. and Irie,R.
TITLE Full-length human cdna
JOURNAL Patent: EP 1440981-A 698 28-JUL-2004;
RESEARCH Association for Biotechnology (JP)
FEATURES
    source
        1..2852
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Query Match 17.3%; Score 722.2; DB 6; Length 2852;
Best Local Similarity 84.8%; Pred. No. 4.1e-184;
Matches 822; Conservative 0; Mismatches 143; Indels 4; Gaps 1;

QY 10 GCTGCTCCGATGCTCCAGAGCGGCATGGGCGCCGCACTGGTGGGACCACTGCGGGC 69
Db 134 GCTGCTCCAAATGCCCGGAGTGGCCATGGGCGCCGCACTGGTGGGACCACTGCGGGC 193
QY 70 TGGCAGTTCGGAGGTGATGTTGGTGGGAGCAACTACACTATCGTGCCTGCCATGCGGA 129
Db 194 TGGTAGTCCGGAGTGGACTGGTGGGAGGACAACTACACCATCGTGCCTGATCGCGGA 253
QY 130 GTTCTACAAACAGATCAGCAAGCTGTTGTTTTCATTTTACCTCCCATCTGCATGTCCT 189
Db 254 GTTCTACAAACAGATCAGCAAGCTGTTTTCATTTTACCTCCCATCTGCATGTCCT 313
QY 190 GTTCCGCGAGTACGCAAGCTGTTTCAACAGCGGCATCTACTTATATGAGGCTCCTAGT 249
Db 314 GTTTCGTCAGTATCAACATGCTTCAACAGTGGCATCTACTTAATCTGGACTCTTTGGT 373
QY 250 TGTAGTGGGATTCGATCTGCTACTTCCATGCAACGCTGAGTTTCCTGGGTACAGTCT 309
Db 374 TGTAGTGGGATTCGATCTGCTACTTCCATGCAACCTTGTAGTTTCTTGGTTCAGTGT 433
QY 310 TGATGAATTCGCCATTCCTGTTGGTTCCTGATGTGCTTTGGCCATGTTGTTTCCAGAG 369
Db 434 TGATGAATTCGAGTCTTTCGTTGTTCTGATGTGCTTTGGCCATGTTGTTTCCAGAG 493
QY 370 GTATTTACCAAGATCTTTCCGATGACAGGGGAGGTTCAAGGAGTGGTGTGTCCT 429

```

```

Db 494 GTATCTACCAAGATCTTTCCGAATGACCGGGGTAGTTTCAAGTGTGGTTCAGTGTCT 553
QY 430 GTCTCAATTACAAAGTGTGGCTTTATCAAGCCGCGCATCAACAATATTTCCTGAT 489
Db 554 GTCTCGGTACGAGTGTCCCGCATTTTGTCAAGCTGCCATCAACAACATCTCTCTGAT 613
QY 490 GATTCTGGGACTTCCATGCACATGCGCTGCTGTGTCAGAGCTGAAGAGGTGTGACAATGT 549
Db 614 GACCTTGGGAGTTCCTTGACATGCACTGCTCATCGCAGAGCTAAAGAGGTGTGACAAT 673
QY 550 GCGTGTGTTTAAAGCTGGGCTCTTCTCTGSCCTCTGGTGGACTCTGGCTCTCTTCTGCTG 609
Db 674 GCGTGTGTTTAAAGCTGGGCTCTTCTCTGSCCTCTGGTGGACCTGCGCTGTCTGCTG 733
QY 610 GATCAGCGAACGAGCTTCTGTGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 669
Db 734 GATCAGTGAACGAGCTTCTGTGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 793
QY 670 TGTGTGGCATATTCTCATCTGCTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 729
Db 794 CATGTGGCACAATCTCATCTGCTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 853
QY 730 TGATGTGCTCTCAGAGATACCTGAGCAAGTCCAGTTCATCAGATTCTGSCCAGCAGAA 789
Db 854 TGATGTGCTCTCAGAGATTCCTGAGCAAGCTGTCATCAAGTTCTGSCCAGCAGAA 913
QY 790 ATGGGGCTTTTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 849
Db 914 ATGGGGCTTTTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 973
QY 850 CAGATCAGTGTGATGCGCAAGGAGTCCAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 909
Db 974 CAGATCAGTGTGATGCGCAAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1029
QY 910 GCTGGGCTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 969
Db 1030 AGTGGGCTTCTTGTCTAGGAGACAGCAAGGAGTTCGAATAGTTGGGTGTGGGCTA 1089
QY 970 TGTTTAAAA 978
Db 1090 TCTTTTCAA 1098

RESULT 8
AKI23581
LOCUS Homo sapiens cDNA FLJ41587 fis, clone CTONG2020638.
DEFINITION AKI23581
ACCESSION AKI23581.1 GI:34529163
VERSION AKI23581
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
Sugiyama,T., Irie,K., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2852)
AUTHORS Isogai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazuma-Kamatori, Kisarazu, Chiba 292-0818, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology (RAB); cDNA library

```

Mapping information for this clone was provided by Dr. John D.



| Query Match           | 16.9%;       | Score 705;  | DB 9;           | Length 149462; |         |
|-----------------------|--------------|---|-----------------|----------------|---------|
| Best Local Similarity | 84.4%;       | Prod. No. 2.7e-179;   |                 |                |         |
| Matches 818;          | Conservative | 4;  | Mismatches 145; | Indels 6;      | Gaps 2; |
| QY                    | 10           | GCTGCTCGGATGCTCCAGAGCGGCATCGGGCCCGGACACGTCGGTGGGACCACTCGGGC     | 69              |                |         |
| Db                    | 136946       | GCTGCTCCAAATGCCCCAGATGGCCATGGCGCCCGGACACTGGTGGGACCACTCGTCGAGGC  |                 |                |         |
| QY                    | 70           | TGGCAGTTCCGAGGTGGATTTGGTGGGAGGACAACTACACTATCTGTCCTTGCCTATGCCGA  | 129             |                |         |
| Db                    | 136886       | TGGCAGCTCGGAGGTGGACTGGCGCGAGGACAACTACACCATCGTGCCTGTCTGCCGA      | 136887          |                |         |
| QY                    | 130          | GTTCTACACACGATCAGCAAGCTCTTGTTTTCAATTTACCTCCCACTCGATGATGCTTT     | 189             |                |         |
| Db                    | 136826       | GTTCTATAACATGATCAGCAATGTCTTAATTTTCAATTTTACCGCCCATCTCGATGATGCTTT | 136767          |                |         |
| QY                    | 190          | GTTCCGCCAGTACGCAAGCTGTCTTCAACAGCGGCATCTACTTAATATGAGCCTCTCTAGT   | 249             |                |         |
| Db                    | 136766       | GTTTCGTGAGTATGCAACATGCTTCAACAGCGGCATCTACTTAATCTGGA--CTCTTGT     | 136709          |                |         |
| QY                    | 250          | TGTAGTGGGGATTGGAATCTGCTACTTTCATGCAAGCGCTGAGTTTCTGGGTGACAGTCT    | 309             |                |         |
| Db                    | 136708       | TGTAGCGGGAATTGGATCCGCTACTTCCATGCAACCCCTAGTTTCTCTGGGTGACAGTCT    | 136649          |                |         |
| QY                    | 310          | TGATGAATCTGCCATCTCTGTGGGTTCTGATGTGCTTTGGCCATGTGTTTCCACGAG       | 369             |                |         |
| Db                    | 136648       | TGATGAATCTGAGTCTCTTGGGTTCTGAATGTGCTTTCGGTCAATGTGTTCCCGAAG       | 136589          |                |         |
| QY                    | 370          | GTATTTACCAAGATCTTTTCGGAATGACAGGGCGCAGGTTCAAGGCAGTGGTGTGTCTCT    | 429             |                |         |
| Db                    | 136588       | GTATCTACCAAGATCTTTTCGGAATGACAGGGTAGGTTCAAGGTGGTGGTCTGTGTCTCT    | 136529          |                |         |
| QY                    | 430          | GTCTGCAATTACAGCTGCTTGGCGTTTATCAAGCCGCGCATCAACAATATTTCCCTGAT     | 489             |                |         |
| Db                    | 136528       | GTCTGCAATTATGAGCTGCTTGGCATTTGTCAAGCCTGCCATCAACAATCTCTCTGAT      | 136469          |                |         |
| QY                    | 490          | GATTCTGGGACTTCCATGCATCTCGCTGCTTGTTCGACAGCTGAAGAGGTGACAAATGT     | 549             |                |         |
| Db                    | 136468       | GACCTTGGGAGTTCCTTGGGCTGCATCTGTCATCACAGAGCTAAGAGGTGACAAAT        | 136409          |                |         |
| QY                    | 550          | CGGTGTGTTAAGCTGGGCCTCTTCTCGGCCTCTGGTGGACTCTGGCTCTCTCTGCTG       | 609             |                |         |
| Db                    | 136408       | CGGTGTGTTAAGCTGGGCCTCTTCTCGGCCTCTGGTGGACCTTGGCCCTGTTCTGCTG      | 136349          |                |         |
| QY                    | 610          | GATCAGCAGCAAGCCTTCTGAGCTGCTCTCTCTCTTTCACATTCCTTACCTGCACTG       | 669             |                |         |
| Db                    | 136348       | GATCAGTGACCGAGTTCCTGCGAGCTGCTGTCCTTCAACTTCCCTTACCTGCACTG        | 136289          |                |         |
| QY                    | 670          | TGTTGGCATATTCATCTGCTTGTCTGCTTCTGCTGGCTGTGTGCTTCTGCTTCTCT        | 729             |                |         |
| Db                    | 136288       | CATGTGGCACATCTCTCATCTGCTTGTCTGCTTCTGCTGGCTGTGTGCTTCTGCTTCT      | 136229          |                |         |
| QY                    | 730          | TGATGCTGCTCAGAGATACCTGAGCAAGGTCCAGTCAATCAAGATTCCTGGCCAGCGAA     | 789             |                |         |
| Db                    | 136228       | TGATGCTGCTCAGAGATTCCTGAGCAAGGCCCTGTCACTAAATTCCTGGCCAGCGAA       | 136169          |                |         |
| QY                    | 790          | ATGGGCTTTTATGGTGTCCCTTATGTGTCCCTTCTGTGTGCCCAAGAGAGTGCCTAGT      | 849             |                |         |
| Db                    | 136168       | ATGGGCTTTTATGGTGTCCCTTATGTGTCCCTTCTGTGTGCCCAAGAGATCATGAT        | 136109          |                |         |
| QY                    | 850          | CAAGATCAGGTGATGGCAAGCGATGACCGATCTCTCTCTCTCTCTCTCTCTCTCTCT       | 909             |                |         |
| Db                    | 136108       | CAAGACCACGTGATGGCAAGATGGTGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT      | 136053          |                |         |
| QY                    | 910          | GCTGGGCTTCTGTTGCTAGCAAGATGAGTCTGAGGGGTTGAGGAATGGTGTGTGTGG       | 969             |                |         |
| Db                    | 136052       | AGTGGGCTTCTGTTGCTAGGAAGACAGCCAAAGGAGTTCAAATAGTTGGGAGTGGGCTA     | 135993          |                |         |
| QY                    | 970          | TGTTTAAAA   | 978             |                |         |
| Db                    | 135992       | TCATTTCAA   | 135984          |                |         |

| RESULT 10             | AX206799   | AX206799                          | AX206799 | 822 bp | DNA | linear | PAT 30-AUG-2001 |
|-----------------------|--|-----------------------------------|----------|--------|-----|--------|-----------------|
| LOCUS                 | AX206799   | Sequence 6 from Patent WO0155408. |          |        |     |        |                 |
| DEFINITION            | AX206799   | Sequence 6 from Patent WO0155408. |          |        |     |        |                 |
| ACCESSION             | AX206799   | Sequence 6 from Patent WO0155408. |          |        |     |        |                 |
| VERSION               | AX206799.1   | GI:15394643                       |          |        |     |        |                 |
| KEYWORDS              |  |                                   |          |        |     |        |                 |
| SOURCE                |  |                                   |          |        |     |        |                 |
| ORGANISM              |  |                                   |          |        |     |        |                 |
| REFERENCE             |  |                                   |          |        |     |        |                 |
| AUTHORS               |  |                                   |          |        |     |        |                 |
| TITLE                 |  |                                   |          |        |     |        |                 |
| JOURNAL               |  |                                   |          |        |     |        |                 |
| FEATURES              |  |                                   |          |        |     |        |                 |
| source                |  |                                   |          |        |     |        |                 |
| ORIGIN                |  |                                   |          |        |     |        |                 |
| Query Match           |  |                                   |          |        |     |        |                 |
| Best Local Similarity |  |                                   |          |        |     |        |                 |
| Matches               |  |                                   |          |        |     |        |                 |
| 725; Conservative     |  |                                   |          |        |     |        |                 |
| 0; Mismatches         |  |                                   |          |        |     |        |                 |
| 97; Indels            |  |                                   |          |        |     |        |                 |
| 3; Gaps               |  |                                   |          |        |     |        |                 |
| 35                    | ATGGGCGCCCGCAGCTGGTGGGACCACTGCGGGCTGGCAGTTCGGAGGTGGATTGGTGC    | 94                                |          |        |     |        |                 |
| 1                     | ATGGGCGCCCGCAGCTGGTGGGACCACTGCGGGCTGGCAGTTCGGAGGTGGATTGGTGC    | 60                                |          |        |     |        |                 |
| 95                    | GAGGACAACTACACTATCGTGCCTCGCATTCGCGAGTTCCTACAAACGATCAGCAACGTC   | 154                               |          |        |     |        |                 |
| 61                    | GAGGACAACTACACCATCGTGCCTCGTGCCTCGCGAGTTCCTACAAACGATCAGCAACGTC  | 120                               |          |        |     |        |                 |
| 155                   | TTGTTTTCATTTTAACCTCCCATCTGCATGTGCTTGTTCGCCAGTACGCAACGTCGCTTC   | 214                               |          |        |     |        |                 |
| 121                   | TTATTTTTCATTTTAACGCCCATCTGCATGTGCTTGTTCGTCAGTATGCAACATGCTTC    | 180                               |          |        |     |        |                 |
| 215                   | AACAGCGGCATCTACTTAATATGACGCTCCTAGTTGTAGTGGGATTGGATCTGCTAC      | 274                               |          |        |     |        |                 |
| 181                   | AACAGCGGCATCTACTTAATCTG---GCTCTGGTTGTAGCGGGAATTGGATCCGCTAC     | 237                               |          |        |     |        |                 |
| 275                   | TTCCATGCAACGCTGAGTTTCCCTGGGTTCAGATGCTTTGATGAATCTGCCATTCCTGGGTT | 334                               |          |        |     |        |                 |
| 238                   | TTCCATGCAACGCTGAGTTTCCCTGGGTTCAGATGCTTTGATGAATCTGCCATTCCTGGGTT | 297                               |          |        |     |        |                 |
| 335                   | CTGATGTGTGCTTTGGCCATGTGGTTTCCAGGAGGTATTACCAAGATCTTTTCGGAAT     | 394                               |          |        |     |        |                 |
| 298                   | CTGATGTGTGCTTTGGGTTCATGTGTTCCCGAAGGTATCTACCAAGATCTTTTCGGAAT    | 357                               |          |        |     |        |                 |
| 395                   | GACAGGGCAGGTTCAAGGAGTGTGTGCTCTCTGCAATTAACAGTGTCTGGCG           | 454                               |          |        |     |        |                 |
| 358                   | GACAGGAGTGTGTTCAAGGTGGTGTCTGTCTCTGCAATTAACAGTGTCTGGCA          | 417                               |          |        |     |        |                 |
| 455                   | TTTATCAAGCCGCCCATCAACAATATTTCCCTGATGATTCCTGGAGCTTCCATGCACTCG   | 514                               |          |        |     |        |                 |
| 418                   | TTTGTCAAGCTGCCATCAACAATCTCTGATGACCTGGAGTTCCTGGCGTGA            | 477                               |          |        |     |        |                 |
| 515                   | CTGCTTTGTTCAGAGCTGAAGAGGTGCAATGTGCGTGTGTTTAAGCTGGGCCCTCTTC     | 574                               |          |        |     |        |                 |
| 478                   | CTGCTCATCACAGCTAAAGAGGTGTGACCAATCGGTGTGTTTAAGCTGGGCCCTCTTC     | 537                               |          |        |     |        |                 |
| 575                   | TCTGGCCTCTGGTGGACTCTGGCTCTCTTCTGCTGGATCAGGACCAAGCCTTCTGTGAG    | 634                               |          |        |     |        |                 |
| 538                   | TCGGGCTCTGGTGGACCTGGGCCCTGTTCTGCTGGATCAGTACGCGAGCTTTCTGCGAG    | 597                               |          |        |     |        |                 |
| 635                   | CTGCTCTCTCTCTTTCACCTTCCCTACCTGCACTGTGTCGCAATCTCATCTGCCCTT      | 694                               |          |        |     |        |                 |
| 598                   | CTGCTGTATCTCTTCAACTTCCCTACCTGCACTGCAATGTGGCAATCTCATCTGCCCTT    | 657                               |          |        |     |        |                 |
| 695                   | GCTTCGTACTGGGCTGTGTGCTTCGCTACTTGTGATGCTGCTCCTCAGAGATACCTGAG    | 754                               |          |        |     |        |                 |



```
Db 658 GGTGCTACTGCTGCTGTGTATGCTTTGGCTACTTTGATGCTGCTCCCTCAGAGATTCCTGAG 7117
QY 755 CAAGGTCAGTCATCAGATTTCTGGCCAGCAGGAAATGGCTTTTATTTGTTGCTCCCTTAT 814
Db 718 CAAGGCTGTGTCATCAAAATTCGGCCAGCAGGAAATGGCTTTTATTTGTTGCTCCCTAT 777
QY 815 GTGTCCCTCTGTGTGTCGCCCAAGAAGTGCCTCAGTCAAGATCAG 859
Db 778 GTGTCCCTCTGTGTGTCGCCCAAGAATCATCATGTCAGGACCAG 822

RESULT 11
G76816 531 bp DNA linear STS 06-SEP-2002
LOCUS S208P6606FG12.T0 129S1/SvImJ Mus musculus STS genomic, sequence
DEFINITION tagged site.
ACCESSION G76816
VERSION G76816.1 GI:22727572
KEYWORDS STS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 531)
Wade, C.
Polymorphism Structure in the Mouse
Unpublished (2002)

REFERENCE
AUTHORS Contact: Kerstin Lindblad-Toh
TITLE Whitehead Institute for Biomedical Research, Center for Genome
JOURNAL Research
COMMENT 320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersli@genome.wi.mit.edu
Primer A: None
Primer B: None
STS size: 531
Protocol:
WGS-discovery: Paired-end low-coverage whole genome shotgun reads
were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
reads were placed uniquely on the MGSV3 C57BL/6J assembly and SNP
detection was carried out by SSAHA-SNP. 225,000 reads were
annotated
as STSS and 81,000 SNPs were annotated with alleles from C57BL/6J
and the strain from which the particular read came. The validation
rate for these SNPs was estimated at approximately 98%.

FEATURES
source Location/Qualifiers
1..531
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="129S1/SvImJ"
/db_xref="taxon:10090"
/map_pos=4 20-461 84241674-84242112"
/clone_lib="129S1/SvImJ"
<1..5531

STS
ORIGIN
Query Match 12.2%; Score 510.8; DB 11; Length 531;
Best Local Similarity 99.1%; Pred. No. 8.4e-127;
Matches 525; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

QY 792 GGGCTTTTATTTGTTGTCCTTATGTGTCCTTCTGTGTGCCCAAGAAGTCGCGAGTCA 851
Db 1 GGGCTTTTATTTGTTGTCCTTATGTGTCCTTCTGTGTGCCCAAGAAGTCGCGAGTCA 60
QY 852 AGATCAGTGATGCGAAGGAGTGACAGCTTCTCTACTTACTTCTATTCAGTGCAGGC 911
Db 61 AGATCAGTGATGCGAAGGAGTGACAGCTTCTCTACTTACTTCTATTCAGTGCAGGC 120
QY 912 TGGGCTTCGTTGCTAGCAAGATGGCTGAGGGGGTTGAGGAATTTGTTGTTGGGTG 971
Db 121 TGGGCTTCGTTGCTAGCAAGATGGCTGAGGGGGTTGAGGAATTTGTTGTTGGGTG 180
```

```
QY 972 TTTAAATTTCTGCTCCTTTGTGATCTAACTGACCACTGTCCTGACCTCCCTAGGTTAA 1031
Db 181 TTTAAATTTCTGCTCCTTTGTGATCTAACTGACCACTGTCCTGACCTCCCTAGGTTAA 240
QY 1032 GTAGAGGGCTCAGGGACATAAGGTGCTTCCTCAGTATCCTTTCCAGGACATACGGGCTTG 1091
Db 241 GTAGAGGGCTCAGGGACATAAGGTGCTTCCTCAGTATCCTTTCCAGGACATACGGGCTTG 300
QY 1092 CTGGGTTATGTCATATGACATCAACAGAGTAGTTCCTTTGGGAGACCTTAGGGCAACCCA 1151
Db 301 CTGGGTTATGTCATATGACATCAACAGAGTAGTTCCTTTGGGAGACCTTAGGGCAACCCA 360
QY 1152 AGTTTCTTGTGCGAGAGGTAGCTGCTGT-----TTCCACCAATTCAGACTCTCAACCC 1208
Db 361 AGTTTCTTGTGCGAGAGGTAGCTGCTGTAGATCCCACTATTCAGACTCTCAACCC 420
QY 1209 CCTGATAAACAAGACCTTCTGATTGTTGATGAAGGTTCCAGAACTTTTCATTTGCGG 1268
Db 421 CCTGATAAACAAGACCTTCTGATTGTTGATGAAGGTTCCAGAACTTTTCATTTGCGG 480
QY 1269 GGAGAAACTGCTCCTTCAACAAAAACCAAGTGGGCAAAACACGTTGGGGGG 1318
Db 481 GGAGAAACTGCTCCTTCAACAAAAACCAAGTGGGCAAAACACGTTGGGGGG 530

RESULT 12
AX400443/c
LOCUS AX400443 595 bp DNA linear PAT 06-JUN-2002
DEFINITION Sequence 119 from Patent WO0210453.
ACCESSION AX400443
VERSION AX400443.1 GI:21336623
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1
REFERENCE Mendrick, D., Porter, M.W., Johnson, K.R., Castle, A.L. and
AUTHORS Elashoff, M.R.
TITLE Molecular toxicology modeling
JOURNAL Patent: WO 0210453-A 119 07-FEB-2002;
Gene Logic, Inc. (US)
FEATURES Location/Qualifiers
source 1..595
/organism="Rattus norvegicus"
/mol_type="unassigned DNA"
/db_xref="taxon:10116"
/notes="EMBL/GenBank Accession No. AA848826"

ORIGIN
Query Match 9.6%; Score 400.4; DB 6; Length 595;
Best Local Similarity 87.9%; Pred. No. 7.8e-97;
Matches 531; Conservative 0; Mismatches 56; Indels 17; Gaps 8;

QY 3557 ATTCTATTGTTGGATTTCTCAGAGTCTGAATCTTCCCTCAGAGTCTCTCTCTCCAA 3616
Db 595 ATTCTATTGTTGGATTTCTCAGAGTCTGAATCTTCCCTCAGAGTCTCTCTCTCTCTCT 536
QY 3617 CCCTACCCAGAGCCACACTGGGATTTGAACCTTC-----TCTGACTCTCTCTCTCTCT 3672
Db 535 CCCTACCCAGAGTTC---CACTGGGATTTGAACCTTCAAATCTCGCTCTCTCTCTCTCT 479
QY 3673 AGGTCTGACAACTAAATGGTCTCTGGGACACCCAGCTAGGGCTTCCCAACTCTCTT-AT 3731
Db 478 AGGTCTGACAACTAAACGCTCTCTGGGACACCTAACTAGGGCTTCCCGAGCTCTCTCT 419
QY 3732 CCAGCTGAACCTTGATTTCTTCCAAACACAGGGCTTGACCTGGGTGCTGTTGGTCCACTGG 3791
Db 418 CCAGCCAAACTTGGATTTCTTCCAAACACAGGGCTTGACCTGGGTGCTGTTGGTCCACTGC 359
QY 3792 CCAACAACACATCTTTGGCCAGATTGGGATTTCAATAGATTTTATAGACATATTATCTCC 3851
```

```
Db 358 CCAATGACACATCTTTGACAGATTGGATTCTCAATAGATTTTATAGACATTTCTCC 299
QY 3852 CACAGACTTT--AAACATGGCTGTGTCTTTCCATACATCCGCTCAGATTAAACT 3909
Db 298 CACAGACTTTAAACAAATGGCGTGTCTTTTCCACACA--AGTCAGATTAAACT 242
QY 3910 ATTATTATACCAAGCAATTAACCAAGCAATAGAGTACTTTTCAGATATAACTGTGT 3969
Db 241 ATTATTATACCAATTAATTAACCAAGCAATAGAGTATTTCAGATATATGC--CAT 184
QY 3970 TCATACCTTTATGTAGAGTGTCTATGTATATAGCGGTATGTACCTCGCTGAAGTAAAT 4029
Db 183 TTATACCTTTATGTAGAGTGTCTATGTATACAGTGGCATGTACCTCGCTGAGTAACT 124
QY 4030 ACCATAGCTCTGGGAGGATTTACAGACCTTTTGCACTTATGC--TTTTTGTGAACCT 4088
Db 123 AGTCATGTCTGGG--GGATTTACAGATCTTTTGCACTTTATGTCTTTTGTGAACCT 65
QY 4089 GATAACCATGGTCAATATTATAAGCCCAATACTGGCATTTTCTGTGAATAAATGCATAT 4148
Db 64 GATAACCATGTTCAATATTATAAGCCCAATACTGGCATTTTCTGTGAATAAATTTATGT 5
QY 4149 GTAT 4152
Db 4 ATCT 1

RESULT 13
BV053534/c
LOCUS
DEFINITION
S212P6219FA6.T0 CZECHII/Ei Mus musculus STS genomic, sequence
tagged site.
ACCESSION
BV053534
VERSION
BV053534.1 GI:31169329
SOURCE
STS.
MUS musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 784)
Wade, C.M., Kulbokas, E.J. III, Kirby, A.W., Zody, M.C., Mullikin, J.C.,
Lander, E.S., Lindblad-Toh, K., and Daly, M.J.
The mosaic structure of variation in the laboratory mouse genome
Nature 420 (6915), 574-578 (2002)
22354684
12466852
COMMENT
Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersti@genome.wi.mit.edu
Primer A: None
Primer B: None
STS size: 784
Protocol:
WGS-discovery: Paired-end low-coverage whole genome shotgun reads
were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP
detection was carried out by SSAHA-SNP. 225,000 reads were
annotated
as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
and the strain from which the particular read came. The validation
rate for these SNPs was estimated at approximately 98%.
Location/Qualifiers
1..784
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="CZECHII/Ei"
/db_xref="taxon:10090"

FEATURES
source
```

```
STSS
ORIGIN
Query Match 9.2%; Score 382.6; DB 11; Length 784;
Best Local Similarity 99.0%; Pred. No. 5.4e-92;
Matches 385; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3767 ACCTGGGTCTGTGTGTCCTCCACTGGCCAAACAACATCTTTGGCCAGATTGGGATTTCTCA 3826
Db 784 ACCTGGGTCTGTGTGTCCTCCACTGGCCAAACAACATCTTTGGCCAGATTGGGATTTCTCA 725
QY 3827 ATAGATTTTATAGACATTTATCTCCACAGACTTTTAAACATGGCTGTGTCTTTCCATA 3886
Db 724 ATATATTTTATAGACATTTATCTCCACAGACTTTTAAACATGGCTGTGTCTTTCTCTTA 665
QY 3887 CACATCCGGTCAAGTTTAAAGCTATTTTATACCAACAGGAATTTAAACCAAGCAATAGAG 3946
Db 664 CACATCCGGTCAAGTTTAAAGCTATTTTATACCAACAGGAATTTAAACCAAGCAATAGAG 605
QY 3947 TACTTTTCAGATATAAAGCTGTGTCTTTCATACCTTTTATGTAGAGTGTGTATATAGGCGGTA 4006
Db 604 TACTTTTCAGATATAAAGCTGTGTCTTTCATACCTTTTATGTAGAGTGTGTATATAGGCGGTA 545
QY 4007 TGTACCTCGCTGAAGTAAATATTAACCATAGCTCTGGGAGGATTTACAGACCTTTTGCAC 4066
Db 544 TGTACCTCGCTGAAGTAAATATTAACCATAGCTCTGGGAGGATTTACAGACCTTTTGCAC 485
QY 4067 TTTATGCTTTTTTTGTGAACCTCTGATAACCATGGTCAATATTAAGCCCAATAAATCTGGCAT 4126
Db 484 TTTATGCTTTTTTTGTGAACCTCTGATAACCATGGTCAATATTAAGCCCAATAAATCTGGCAT 425
QY 4127 TTCTGTGAATAAATGCATATGTATCTA 4155
Db 424 TTCTGTGAATAAATGCATATGTATCTA 396

RESULT 14
CQ736905
LOCUS
DEFINITION
Sequence 22839 from Patent WO02068579.
ACCESSION
CQ736905
VERSION
CQ736905.1 GI:42333763
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
Patent: WO 02068579-A 22839 06-SEP-2002;
PE Corporation (NY) (US)
JOURNAL
Location/Qualifiers
1..315
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

FEATURES
source
ORIGIN
Query Match 6.3%; Score 261.2; DB 6; Length 315;
Best Local Similarity 89.5%; Pred. No. 4.5e-59;
Matches 281; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 549 TGCGTGTGTTTAAAGCTGGGCTCTTCTCTGGCTCTGTGGACTCTGGCTCTCTTCTGTCT 608
Db 2 TGCGTGTGTTTAAAGCTGGGCTCTTCTCTGGGCTCTGTGGACTCTGGCTCTCTTCTGTCT 61
QY 609 GGATCAGCGACCAAGCCCTTCTGTAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 668
```



*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 20:48:16 ; Search time 1770.81 Seconds  
(without alignments)  
12376.463 Million cell updates/sec

Title: US-10-017-410-1  
Perfect score: 4175  
Sequence: 1 ggcacagggtctccgat.....aaaaaaaaaaaaaaaaaac 4175

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues  
Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_23Sep04.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 4175  | 100.0       | 4175   | 6     | ABSS54148          |
| 2          | 723.8 | 17.3        | 4202   | 8     | ACC50961 Human bla |
| 3          | 723.8 | 17.3        | 4212   | 11    | Adn38833 Cancer/an |
| 4          | 684.8 | 15.9        | 828    | 6     | ABSS54149          |
| 5          | 653.8 | 15.7        | 822    | 4     | Aah48641 Human cer |
| 6          | 424.2 | 10.2        | 1623   | 10    | Adi21948 Novel hum |
| 7          | 400.4 | 9.6         | 595    | 6     | Abk62212 Rat seque |
| 8          | 400.4 | 9.6         | 595    | 10    | ABSS5226           |
| 9          | 323.6 | 7.8         | 1527   | 10    | ADL06640           |
| 10         | 323.6 | 7.8         | 1527   | 10    | ADL06642 Human 3T3 |
| 11         | 294.8 | 7.1         | 487    | 10    | Adi21468           |
| 12         | 176.8 | 4.2         | 1215   | 4     | AAK51566           |
| 13         | 176.8 | 4.2         | 1631   | 4     | AAK52550 Human pol |
| 14         | 174.8 | 4.2         | 792    | 4     | Aah48639 Human cer |
| 15         | 160.6 | 3.8         | 537    | 4     | ABL24043           |
| 16         | 125.8 | 3.0         | 1473   | 4     | AAH33545           |
| 17         | 108   | 2.6         | 2143   | 5     | ABV24475 Human col |
| 18         | 108   | 2.6         | 2143   | 5     | ABV24620 Human pro |
| 19         | 108   | 2.6         | 2143   | 5     | ABV24928 Human pro |
| 20         | 92.6  | 2.2         | 4758   | 4     | ABL24042 Drosophil |
| 21         | 77.8  | 1.9         | 4071   | 4     | ABL30432 Drosophil |

|   |    |      |     |        |    |            |
|---|----|------|-----|--------|----|------------|
| c | 22 | 70.6 | 1.7 | 692    | 5  | ABV28857   |
| c | 23 | 70.6 | 1.7 | 692    | 5  | ABV23021   |
| c | 24 | 61.4 | 1.5 | 110000 | 8  | ABX16390_3 |
| c | 25 | 59   | 1.4 | 2000   | 8  | ADA71938   |
| c | 26 | 58.2 | 1.4 | 35425  | 10 | ADC87616   |
| c | 27 | 54.6 | 1.3 | 14658  | 12 | ADM80713   |
| c | 28 | 52.6 | 1.3 | 19619  | 6  | ABK47192   |
| c | 29 | 49.8 | 1.2 | 2000   | 8  | ADA71938   |
| c | 30 | 49.4 | 1.2 | 7516   | 6  | ABL32061   |
| c | 31 | 48.6 | 1.2 | 50295  | 9  | ADA02993   |
| c | 32 | 48.6 | 1.2 | 50295  | 10 | ADB72731   |
| c | 33 | 48.6 | 1.2 | 50295  | 10 | ADC85473   |
| c | 34 | 48.6 | 1.2 | 50295  | 12 | ADM74588   |
| c | 35 | 47.6 | 1.1 | 52479  | 9  | ADA02795   |
| c | 36 | 47.6 | 1.1 | 52479  | 10 | ADB72533   |
| c | 37 | 47.6 | 1.1 | 52479  | 10 | ADC85275   |
| c | 38 | 47.6 | 1.1 | 52479  | 12 | ADM74390   |
| c | 39 | 47.4 | 1.1 | 215980 | 6  | AAL38337   |
| c | 40 | 47   | 1.1 | 5734   | 10 | ADC85382   |
| c | 41 | 47   | 1.1 | 67191  | 9  | ADA02903   |
| c | 42 | 47   | 1.1 | 67191  | 10 | ADB72641   |
| c | 43 | 47   | 1.1 | 67191  | 11 | ADL95495   |
| c | 44 | 47   | 1.1 | 67191  | 12 | ADM74498   |
| c | 45 | 46.6 | 1.1 | 171936 | 6  | ABS56565   |

## ALIGNMENTS

RESULT 1  
ABS54148  
ID ABS54148 standard; cDNA; 4175 BP.  
XX AC ABS54148;  
XX DT 25-NOV-2002 (first entry)  
XX DE Mouse cDNA encoding a liver tumour marker protein, CRG-L1.  
XX KW Mouse; ss; gene; liver cancer; liver tumour; CRG-L1;  
XX KW hepatocellular cancer.  
XX OS Mus musculus.  
XX FH Key Location/Qualifiers  
FT CDS 35..862  
FT /\*tag= a  
FT /product= "Liver tumour marker protein, CRG-L1"  
XX US2002115094-A1.  
XX PD 22-AUG-2002.  
XX PF 14-DEC-2001; 2001US-00017410.  
XX PR 14-DEC-2000; 2000US-0255674P.  
XX (FARN/) FARNHAM P J.  
XX (GRAV/) GRAVEEL C R.  
XX FI Farnham PJ, Graveel CR;  
XX WPI; 2002-706409/76.  
XX P-PSDB; ABG32880.  
XX PT Novel polypeptide designated as CRG-L1, useful as diagnostic marker for  
XX liver cancer, is differentially expressed in liver tumours relative to  
XX normal liver tissues.  
XX PS Claim 2; Page 4-7; 11pp; English.  
XX CC The invention relates to a polypeptide designated as CRG-L1, which is  
XX differentially expressed in liver tumours relative to normal expression

CC in normal liver tissues, designated CRG-L1. Also included are the  
 CC encoding polynucleotides (in the case of the human sequence, mapping to  
 CC chromosome 9p), expression constructs, host cells, anti-CRG-L1  
 CC antibodies, identifying modulators of CRG-L1, and the use of the CRG-L1  
 CC sequence in the diagnosis of hepatocellular cancer in tumour cells from a  
 CC liver of a human or non-human animal. The CRG-L1 protein and  
 CC polynucleotide are useful as diagnostic markers for a liver cancer in  
 CC humans and non-human animals, and as a system for assessing putative  
 CC therapeutic agents. The present sequence encodes mouse CRG-L1  
 XX

|                       |                |  |          |             |  |
|-----------------------|----------------|--|----------|-------------|--|
| Query Match           | 100.0%         | Score 4175   | DB 6     | Length 4175 |  |
| Best Local Similarity | 100.0%         | Pred. No. 0  |          |             |  |
| Matches 4175          | Conservative 0 | Mismatches 0   | Indels 0 | Gaps 0      |  |
| QY                    | 1              | GGCAGAGGCTGCTCCGATGCTCCAGAGCGGCATGGCGCCCGCACTGGTGGGACCA      | 60       |             |  |
| DB                    | 1              | GGCAGAGGCTGCTCCGATGCTCCAGAGCGGCATGGCGCCCGCACTGGTGGGACCA      | 60       |             |  |
| QY                    | 61             | CTTGCGGGCTGGCAGTTCGGAGTGGATGGTGGGAGGACCACTACACTATGCTGCTGC    | 120      |             |  |
| DB                    | 61             | CTTGCGGGCTGGCAGTTCGGAGTGGATGGTGGGAGGACCACTACACTATGCTGCTGC    | 120      |             |  |
| QY                    | 121            | CATTGCGGAGTCTCAACACAGATCAGCAAGCTCTTGTGTTTTCATTTTACCTCCCATCTG | 180      |             |  |
| DB                    | 121            | CATTGCGGAGTCTCAACACAGATCAGCAAGCTCTTGTGTTTTCATTTTACCTCCCATCTG | 180      |             |  |
| QY                    | 181            | CATTGCTTTGTTCCGCGCAGTACGCAACGCTCTTCAACAGCGGCATCTACTTAATGGAC  | 240      |             |  |
| DB                    | 181            | CATTGCTTTGTTCCGCGCAGTACGCAACGCTCTTCAACAGCGGCATCTACTTAATGGAC  | 240      |             |  |
| QY                    | 241            | GCTCCTAGTGTAGTGGGAGTGGATCTGTCTACTTCCATGCAACGCTGAGTTTCTGGG    | 300      |             |  |
| DB                    | 241            | GCTCCTAGTGTAGTGGGAGTGGATCTGTCTACTTCCATGCAACGCTGAGTTTCTGGG    | 300      |             |  |
| QY                    | 301            | TCAGATGCTTGATGAATCTGCCATCTCTGGGTTCTGATGTGTCTGGCCCATGGTT      | 360      |             |  |
| DB                    | 301            | TCAGATGCTTGATGAATCTGCCATCTCTGGGTTCTGATGTGTCTGGCCCATGGTT      | 360      |             |  |
| QY                    | 361            | TCCAGGAGTATTTACCAAGATCTTTCCGAATGACAGGGCAGGTTCAAGSCAGTGT      | 420      |             |  |
| DB                    | 361            | TCCAGGAGTATTTACCAAGATCTTTCCGAATGACAGGGCAGGTTCAAGSCAGTGT      | 420      |             |  |
| QY                    | 421            | GTGTGCTCTGTGCAATTTACCAAGTCTTGGGTTTATCAAGCCGCCCATCAACAATAT    | 480      |             |  |
| DB                    | 421            | GTGTGCTCTGTGCAATTTACCAAGTCTTGGGTTTATCAAGCCGCCCATCAACAATAT    | 480      |             |  |
| QY                    | 481            | TTCCCTGATGATCTGGGACTTCCATGCACTGGCTGCTTGTGTCAGAGCTGAAGAGTG    | 540      |             |  |
| DB                    | 481            | TTCCCTGATGATCTGGGACTTCCATGCACTGGCTGCTTGTGTCAGAGCTGAAGAGTG    | 540      |             |  |
| QY                    | 541            | TGACAAATGTCGCTGTTTAAAGCTGGGCTCTTCTCTGGCCCTCTGGTGGACTCTGGCTCT | 600      |             |  |
| DB                    | 541            | TGACAAATGTCGCTGTTTAAAGCTGGGCTCTTCTCTGGCCCTCTGGTGGACTCTGGCTCT | 600      |             |  |
| QY                    | 601            | CTTCTGCTGGATCAGGACCAAGCTCTGTGAGCTGCTCTCTCTCTCTCTCTCTCTCT     | 660      |             |  |
| DB                    | 601            | CTTCTGCTGGATCAGGACCAAGCTCTGTGAGCTGCTCTCTCTCTCTCTCTCTCTCT     | 660      |             |  |
| QY                    | 661            | CCTGCACTGTGTGGGATATTTCTCATCTGCTGCTTGTCTGCTGCTGCTGCTGCTGCT    | 720      |             |  |
| DB                    | 661            | CCTGCACTGTGTGGGATATTTCTCATCTGCTGCTTGTCTGCTGCTGCTGCTGCTGCT    | 720      |             |  |
| QY                    | 721            | CGCTACTTTTGTGCTGCTCAGAGATACCTGAGCAAGGTTCAGTCACTGAGTTCGGGC    | 780      |             |  |
| DB                    | 721            | CGCTACTTTTGTGCTGCTCAGAGATACCTGAGCAAGGTTCAGTCACTGAGTTCGGGC    | 780      |             |  |
| QY                    | 781            | CAGCGAGAAATGGGCTTTTATTTGGTCTTCTGCTTATGCTCTCTGCTGCTGCTGCTG    | 840      |             |  |
| DB                    | 781            | CAGCGAGAAATGGGCTTTTATTTGGTCTTCTGCTTATGCTCTCTGCTGCTGCTGCTG    | 840      |             |  |
| QY                    | 841            | GTGCGCAGTCAAGATCACGTCATGGGAGGCAAGGCAAGTCTCTCTACTTCTTCTATT    | 900      |             |  |

|    |      |   |      |  |  |
|----|------|---|------|--|--|
| DB | 841  | GTGCGCAGTCAAGATCACGTCATGGGAGGCAAGTCTCTCTACTTCTTCTATT          | 900  |  |  |
| QY | 901  | CGAGTGGCGCTGGGCTTCGTTTGTAGCAAGATGGCTGAGGGGTGTGAGGAATGGTG      | 960  |  |  |
| DB | 901  | CGAGTGGCGCTGGGCTTCGTTTGTAGCAAGATGGCTGAGGGGTGTGAGGAATGGTG      | 960  |  |  |
| QY | 961  | TGTTGTGGGTGTTTAAATTTCTGCTCTTGTGATCTAACTGGACCACTGTGCTGACCT     | 1020 |  |  |
| DB | 961  | TGTTGTGGGTGTTTAAATTTCTGCTCTTGTGATCTAACTGGACCACTGTGCTGACCT     | 1020 |  |  |
| QY | 1021 | CCTTAGTGAAGTAGAGGCTCAGGACATAGGTGCTCTCTCCTCAGTATCTCTTCAGGC     | 1080 |  |  |
| DB | 1021 | CCTTAGTGAAGTAGAGGCTCAGGACATAGGTGCTCTCTCCTCAGTATCTCTTCAGGC     | 1080 |  |  |
| QY | 1081 | ATACGGGCTTGTGGGTATGTCATTAATGACATCAACAGAGTAGTTCTTTGGGAGACCT    | 1140 |  |  |
| DB | 1081 | ATACGGGCTTGTGGGTATGTCATTAATGACATCAACAGAGTAGTTCTTTGGGAGACCT    | 1140 |  |  |
| QY | 1141 | AGGGCAACCCAAAGTTTCTTGTGGAGAGGTAGTGTGTTTCCACCATTTCCAGACT       | 1200 |  |  |
| DB | 1141 | AGGGCAACCCAAAGTTTCTTGTGGAGAGGTAGTGTGTTTCCACCATTTCCAGACT       | 1200 |  |  |
| QY | 1201 | CTCAACCCCTGATATAAACAAGACCTTCTGATTTGGTGTGTAAGAGTTCCAGAACTTTTCA | 1260 |  |  |
| DB | 1201 | CTCAACCCCTGATATAAACAAGACCTTCTGATTTGGTGTGTAAGAGTTCCAGAACTTTTCA | 1260 |  |  |
| QY | 1261 | TTTTGCCGGAGAACTGCTTCAACAAAAACAAGTGGGCAAAAACGCTGTGGGGGTG       | 1320 |  |  |
| DB | 1261 | TTTTGCCGGAGAACTGCTTCAACAAAAACAAGTGGGCAAAAACGCTGTGGGGGTG       | 1320 |  |  |
| QY | 1321 | TGACTAAGACGGGTGGCTTGTCACTCGCACCTCTTAGCTTTCCAGGTTCTCCACGTT     | 1380 |  |  |
| DB | 1321 | TGACTAAGACGGGTGGCTTGTCACTCGCACCTCTTAGCTTTCCAGGTTCTCCACGTT     | 1380 |  |  |
| QY | 1381 | TGTGATCTCTCCACTACCTCTGCTGAGAGATGGACCAAGCTTCCAGAGGCAAAAGCTG    | 1440 |  |  |
| DB | 1381 | TGTGATCTCTCCACTACCTCTGCTGAGAGATGGACCAAGCTTCCAGAGGCAAAAGCTG    | 1440 |  |  |
| QY | 1441 | GCAACACCTCTATGTCAAAAAGCTACCTCTTTTAAAGCACACATTATACAGACTAT      | 1500 |  |  |
| DB | 1441 | GCAACACCTCTATGTCAAAAAGCTACCTCTTTTAAAGCACACATTATACAGACTAT      | 1500 |  |  |
| QY | 1501 | TTCTGCTCTTCCAGAGTGCAGAGCTCAGACCAAGCTTCCAGAGAACTCTTCAAGTTATG   | 1560 |  |  |
| DB | 1501 | TTCTGCTCTTCCAGAGTGCAGAGCTCAGACCAAGCTTCCAGAGAACTCTTCAAGTTATG   | 1560 |  |  |
| QY | 1561 | GATTCACACACAGCCCTCTTGTGACCTCTGTGAGAGCAAGGCTCGTTGTAATCGCG      | 1620 |  |  |
| DB | 1561 | GATTCACACACAGCCCTCTTGTGACCTCTGTGAGAGCAAGGCTCGTTGTAATCGCG      | 1620 |  |  |
| QY | 1621 | GAGGAGCTCTGGAGCTTGAAGTGGAGGCTGTGGAAGATTTTGGAAAGCTTTTCTTA      | 1680 |  |  |
| DB | 1621 | GAGGAGCTCTGGAGCTTGAAGTGGAGGCTGTGGAAGATTTTGGAAAGCTTTTCTTA      | 1680 |  |  |
| QY | 1681 | TTGGAGCTTTTGGTGTGCTGCTTACCTCAAGATTCAGACCCCTCTCTGCTGTTT        | 1740 |  |  |
| DB | 1681 | TTGGAGCTTTTGGTGTGCTGCTTACCTCAAGATTCAGACCCCTCTCTGCTGTTT        | 1740 |  |  |
| QY | 1741 | AGGGATATGTCCTCTGAGTGAAGTACCTCTTAGGATGTGGGCTCCGGCTTTTGT        | 1800 |  |  |
| DB | 1741 | AGGGATATGTCCTCTGAGTGAAGTACCTCTTAGGATGTGGGCTCCGGCTTTTGT        | 1800 |  |  |
| QY | 1801 | TTTCATAATACCTGGTGTAACTGGTTTCTCAGAGCACTTGTCTCTTGTGCTGGGC       | 1860 |  |  |
| DB | 1801 | TTTCATAATACCTGGTGTAACTGGTTTCTCAGAGCACTTGTCTCTTGTGCTGGGC       | 1860 |  |  |
| QY | 1861 | GGTCACTACACTGTGATTTGGCTGGGATCCAGAGGAGAAAGGGCAGACTCTGACA       | 1920 |  |  |
| DB | 1861 | GGTCACTACACTGTGATTTGGCTGGGATCCAGAGGAGAAAGGGCAGACTCTGACA       | 1920 |  |  |
| QY | 1921 | TGAGTGTCTCTCACCTGCACTCTACTGCTTCACTGGGCTCTCTTCCAGGGTATGAAGTGG  | 1980 |  |  |

Db 1921 TGAATGTCCTCACTGTCATCTTCTTCACTGGGCTCTCTTCAAGGTATGAAGTGG 1980  
 QY 1981 GTATGGGGTATCTCAGGGAATGTTGTAACTCAGGCACCTTCTGCTTTCTGACATTCCTAT 2040  
 Db 1981 GTATGGGGTATCTCAGGGAATGTTGTAACTCAGGCACCTTCTGCTTTCTGACATTCCTAT 2040  
 QY 2041 GTCACTGTGAGACTGCACTCAGGGGACTGGAGTTGGAACCCGTTCTGAGTGGTGTCT 2100  
 Db 2041 GTCACTGTGAGACTGCACTCAGGGGACTGGAGTTGGAACCCGTTCTGAGTGGTGTCT 2100  
 QY 2101 GAGTCTGAAGGAGTGTAGCTGGGGTCTTGGGACCTCTTAGGATCTCTTACTCTGTTTGA 2160  
 Db 2101 GAGTCTGAAGGAGTGTAGCTGGGGTCTTGGGACCTCTTAGGATCTCTTACTCTGTTTGA 2160  
 QY 2161 CTTTCAAGGGTACAAGTGGGAACTGGAATTAAGAGTGTGTTTAAATGAATGACTTCAT 2220  
 Db 2161 CTTTCAAGGGTACAAGTGGGAACTGGAATTAAGAGTGTGTTTAAATGAATGACTTCAT 2220  
 QY 2221 TCTGCTTTTGGGCTTTGGAGTGTAAAGTAACTTGGCTGGGCTATTTCTGTTTATAGT 2280  
 Db 2221 TCTGCTTTTGGGCTTTGGAGTGTAAAGTAACTTGGCTGGGCTATTTCTGTTTATAGT 2280  
 QY 2281 CACAAATATTTATAGAACATGAAGGTGTAATAAGTGTCTTTTATTAATTCATAGCA 2340  
 Db 2281 CACAAATATTTATAGAACATGAAGGTGTAATAAGTGTCTTTTATTAATTCATAGCA 2340  
 QY 2341 TTTACCAACTCTCCAGGTAGCAACACACAGTCAATTTGATGGGGAATTAAGGGCAGG 2400  
 Db 2341 TTTACCAACTCTCCAGGTAGCAACACACAGTCAATTTGATGGGGAATTAAGGGCAGG 2400  
 QY 2401 AATACTTGGTGTGAATGATGATATGTCATTTTACTGTAGCTTAAATATGTTTAA 2460  
 Db 2401 AATACTTGGTGTGAATGATGATATGTCATTTTACTGTAGCTTAAATATGTTTAA 2460  
 QY 2461 TGACTTTGAGTGAATGTTTGTGGCAACTAGGGGTATATGGATCAGAAATTCGGGCTGAG 2520  
 Db 2461 TGACTTTGAGTGAATGTTTGTGGCAACTAGGGGTATATGGATCAGAAATTCGGGCTGAG 2520  
 QY 2521 TCCCTTTGGTGTGAGTCTGAAATTTTCAAGGGCAAAATCAAACTAGCTTGTTCG 2580  
 Db 2521 TCCCTTTGGTGTGAGTCTGAAATTTTCAAGGGCAAAATCAAACTAGCTTGTTCG 2580  
 QY 2581 GTTCATCTTACAGGCTGGGAGTGCCTTATATATTTCTACATAGGACTTATATACA 2640  
 Db 2581 GTTCATCTTACAGGCTGGGAGTGCCTTATATATTTCTACATAGGACTTATATACA 2640  
 QY 2641 TAAGTCTGTATAATGTCCTGAAGATGACACTAGCTGCCTTCACTCGAAGGGTGTCT 2700  
 Db 2641 TAAGTCTGTATAATGTCCTGAAGATGACACTAGCTGCCTTCACTCGAAGGGTGTCT 2700  
 QY 2701 GGGGCTGGAGATGGTTCACAGTTAAGAACCCGTTGTTGCTCTCTCCAGACGCCAG 2760  
 Db 2701 GGGGCTGGAGATGGTTCACAGTTAAGAACCCGTTGTTGCTCTCTCCAGACGCCAG 2760  
 QY 2761 CGACTGTGAGGCTCATAAACAGCAACTCCAGCCCTGGGCACTCTGAGCCCTCTCTGA 2820  
 Db 2761 CGACTGTGAGGCTCATAAACAGCAACTCCAGCCCTGGGCACTCTGAGCCCTCTCTGA 2820  
 QY 2821 CTTCTAAGGGCATTTGGTGCACAGTACATACAGGCGGCAAAATATATACATGTAA 2880  
 Db 2821 CTTCTAAGGGCATTTGGTGCACAGTACATACAGGCGGCAAAATATATACATGTAA 2880  
 QY 2881 GTAATAATGATAGTTAGTGAGCGGCTGAAGAAAGGAGTGTATGATCAAGGTTTA 2940  
 Db 2881 GTAATAATGATAGTTAGTGAGCGGCTGAAGAAAGGAGTGTATGATCAAGGTTTA 2940  
 QY 2941 GTCTGACCTGAGTGCCTTCTTACCGCGATGTTCTCAAGTCAAGGCAAGCCAT 3000  
 Db 2941 GTCTGACCTGAGTGCCTTCTTACCGCGATGTTCTCAAGTCAAGGCAAGCCAT 3000  
 QY 3001 TTTCTTTTATCAGGAAAGGACTTATTTCCCTAGGGCTCTGCTGACATCTCCTAGGAACA 3060  
 Db 3001 TTTCTTTTATCAGGAAAGGACTTATTTCCCTAGGGCTCTGCTGACATCTCCTAGGAACA 3060

QY 3061 GAGATAAAATACCGATGGATGTAATGAATGATGCTTGGTGAAGAGCCGAGTACCTGA 3120  
 Db 3061 GAGATAAAATACCGATGGATGTAATGAATGATGCTTGGTGAAGAGCCGAGTACCTGA 3120  
 QY 3121 CTGACCCAGTGGGCACTTCCACAGCGAAGCCGGTAAACGGATGTCACCTCCGAATC 3180  
 Db 3121 CTGACCCAGTGGGCACTTCCACAGCGAAGCCGGTAAACGGATGTCACCTCCGAATC 3180  
 QY 3181 TTTCTCCCTCTCTGGTAGAGATGACATCGATGATGTCGTGACCACTGGGCCCCGATG 3240  
 Db 3181 TTTCTCCCTCTCTGGTAGAGATGACATCGATGATGTCGTGACCACTGGGCCCCGATG 3240  
 QY 3241 GGTGTGGGCTCTCTTCTTCTTCCGAATCTACTGAGATCTCAGGACAGAGGACCATGA 3300  
 Db 3241 GGTGTGGGCTCTCTTCTTCTTCCGAATCTACTGAGATCTCAGGACAGAGGACCATGA 3300  
 QY 3301 ATGTTACCAAGTGGTCTGCTGCTGAGTATTTTACAGTGTGTTGACCAATTTTGGTTTT 3360  
 Db 3301 ATGTTACCAAGTGGTCTGCTGCTGAGTATTTTACAGTGTGTTGACCAATTTTGGTTTT 3360  
 QY 3361 AGGAGAATTTCTTCTCTTCTAGTGCCTGTGATGCCAAGCCAGCCCTCAGAAAGTTC 3420  
 Db 3361 AGGAGAATTTCTTCTCTTCTAGTGCCTGTGATGCCAAGCCAGCCCTCAGAAAGTTC 3420  
 QY 3421 CTTCTGCTGCTTCCCTTTGTAATGATCCACTGGGGAATGGGTGTCTACCCAGGGA 3480  
 Db 3421 CTTCTGCTGCTTCCCTTTGTAATGATCCACTGGGGAATGGGTGTCTACCCAGGGA 3480  
 QY 3481 AAAGTGCCCTACCCAGTCAAGGCTCAAGGTTGTTGTTGAGGATCAACATGGCTCTGT 3540  
 Db 3481 AAAGTGCCCTACCCAGTCAAGGCTCAAGGTTGTTGTTGAGGATCAACATGGCTCTGT 3540  
 QY 3541 GAAATAACTGCGCACCCATTTCTATTGGTTGGATTTCTCAGGAGTCTGAATCTTCCCTCAG 3600  
 Db 3541 GAAATAACTGCGCACCCATTTCTATTGGTTGGATTTCTCAGGAGTCTGAATCTTCCCTCAG 3600  
 QY 3601 AGTCCCTTTCTCCAAACCCCTACCCAGAGCAACCACTGGGATTTGAACCTTCTCTGACTC 3660  
 Db 3601 AGTCCCTTTCTCCAAACCCCTACCCAGAGCAACCACTGGGATTTGAACCTTCTCTGACTC 3660  
 QY 3661 TCTTCTCCCTCAGGCTGACAACTAATGGTCTCTGGGACACCCAGCTAGGGCTTCCC 3720  
 Db 3661 TCTTCTCCCTCAGGCTGACAACTAATGGTCTCTGGGACACCCAGCTAGGGCTTCCC 3720  
 QY 3721 CAACTCCCTTATCCAGCTGAACCTTGGATTTCTCCCAACCCAGGCTTGAACCTTCTCTGACTC 3780  
 Db 3721 CAACTCCCTTATCCAGCTGAACCTTGGATTTCTCCCAACCCAGGCTTGAACCTTCTCTGACTC 3780  
 QY 3781 GGTCCCACTGGCCAAACACATCTTTGGCCAGATTTGGGATTTCAATAGATTTTATAGA 3840  
 Db 3781 GGTCCCACTGGCCAAACACATCTTTGGCCAGATTTGGGATTTCAATAGATTTTATAGA 3840  
 QY 3841 CATTAATTTCCCAAGACTTTAAACATGGCTTCTTCCCAACCCAGGCTTGAACCTTCTCTGACTC 3900  
 Db 3841 CATTAATTTCCCAAGACTTTAAACATGGCTTCTTCCCAACCCAGGCTTGAACCTTCTCTGACTC 3900  
 QY 3901 TTTTAAACTATTTTATTAACCAAGGAATTTAAACCAAGCAATAGAGTACTTTTCAGATATA 3960  
 Db 3901 TTTTAAACTATTTTATTAACCAAGGAATTTAAACCAAGCAATAGAGTACTTTTCAGATATA 3960  
 QY 3961 AACTGTGTTTCACTATTTATGATAGTGTGCTATGATAGCGGATGTACCCCTGGCTGA 4020  
 Db 3961 AACTGTGTTTCACTATTTATGATAGTGTGCTATGATAGCGGATGTACCCCTGGCTGA 4020  
 QY 4021 AGTAATTTAACCAATAGCTCTGGGAGGATTTTACAGACCTTTTGCACCTTTATGCTTTTTT 4080  
 Db 4021 AGTAATTTAACCAATAGCTCTGGGAGGATTTTACAGACCTTTTGCACCTTTATGCTTTTTT 4080  
 QY 4081 TGAACCTCTGATAACCATGGTCAATATTTAAAGCCCAATTAACCTGGCAATTTTCTGTGAATAAC 4140  
 Db 4081 TGAACCTCTGATAACCATGGTCAATATTTAAAGCCCAATTAACCTGGCAATTTTCTGTGAATAAC 4140



QY 4141 ATGCATATGATCTAATAAAAAAAAAAAAAAAAAAAAAA 4175  
DB 4141 ATGCATATGATCTAATAAAAAAAAAAAAAAAAAAAAAA 4175

RESULT 2  
ACC50961  
ID ACC50961 standard; cDNA; 4202 BP.  
XX  
AC ACC50961;  
XX  
DT 12-JUN-2003 (first entry)  
XX  
DE Human bladder cancer associated cDNA sequence SEQ ID NO:21.  
XX  
KW Human; bladder cancer; cytostatic; gene therapy; vaccine; gene; es.  
XX  
OS Homo sapiens.  
XX  
FN WO200303906-A2.  
XX  
PD 16-JAN-2003.  
XX  
PF 03-JUL-2002; 2002WO-US0211338.  
XX  
PR 03-JUL-2001; 2001US-0302814P.  
PR 03-AUG-2001; 2001US-0310099P.  
PR 08-NOV-2001; 2001US-0343705P.  
PR 13-NOV-2001; 2001US-0350666P.  
PR 12-APR-2002; 2002US-0372246P.  
XX  
PA (BOSB-) EOS BIOTECHNOLOGY INC.  
XX  
PI Mack DH, Aziz N;  
XX  
WP 1; 2003-201532/19.  
DR P-PSDB; ABR48156.  
XX

PT Detecting a bladder cancer-associated transcript in a cell from a  
PT patient, comprises contacting a biological sample from the patient with a  
PT bladder cancer-associated polynucleotide or antibody.  
XX  
PS Claim 6; Page 235; 307pp; English.  
XX  
CC The present invention describes a method for detecting a bladder cancer-  
CC associated transcript in a cell from a patient. The method comprises  
CC contacting a biological sample from the patient with a polynucleotide  
CC that selectively hybridizes to a sequence that is 80 % identical to a  
CC table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059  
CC encode the human bladder cancer-associated proteins given in ABR48146 to  
CC ABR48242). Bladder cancer-associated sequences from the present invention  
CC have cytostatic activities, and can be used in antisense gene therapy and  
CC in vaccine production. The method can be used for detecting a bladder  
CC cancer-associated transcript in a cell from a patient. The method is  
CC useful in diagnosing or treating bladder cancer and in screening for  
CC compounds that modulate bladder cancer, such as hormones or antibodies.  
CC The nucleic acid molecules from the present invention may be used in  
CC various screening and diagnostic methods, and for gene therapy, vaccine  
CC and/or antisense/inhibition applications  
XX  
SQ Sequence 4202 BP; 1045 A; 892 C; 927 G; 1338 T; 0 U; 0 Other;

Query Match 17.3%; Score 723.8; DB 8; Length 4202;  
Best Local Similarity 84.9%; Pred. No. 5.5e-189;  
Matches 823; Conservative 0; Mismatches 142; Indels 4; Gaps 1;  
QY 10 GCTGCTCCGATGCTCCAGAGCGGCATGGGCGCCCGCATGGTGGGACCACTGGCGGC 69  
DB 52 GCTGCTCCATGCTCCCGAGTGGCCATGGGCGCCCGCATGGTGGGACCACTGGCGGC 111  
QY 70 TGGCAGTTCGAGTGGATGGTGGGAGGACCACTACACTATCTGCTGCTATGGCGGA 129  
DB 112 TGGTAGCTCGAGGTGGACTGGTGGGAGGACCACTACACTATCTGCTGCTATGGCGGA 171

QY 130 GTTCTACACACGATCAGCAACGCTCTGTTTTCATTTTACCTCCCATCTGCATGTGCTT 189  
DB 172 GTTCTACACACGATCAGCAACGCTCTGTTTTCATTTTACCTCCCATCTGCATGTGCTT 231  
QY 190 GTTCCGCCAGTACGCAACGCTCTCAACAGCGGCATCTACTTAATATGAGCGCTCTCTAGT 249  
DB 232 GTTTCGTGATGCAACATGCTTCAACAGTGCATCTACTTAATCTGGACTCTTTTGGT 291  
QY 250 TGTAGTGGGATTTGATCTGTCTACTTCCATGCAACGCTGAGTTTCTGGGTGAGATGCT 309  
DB 292 TGTAGTGGGATTTGATCTGTCTACTTCCATGCAACGCTTGTGCTTCTGGGTGAGATGCT 351  
QY 310 TGTAGTGGGATTTGATCTGTCTACTTCCATGCAACGCTGAGTTTCTGGGTGAGATGCT 369  
DB 352 TGTAGTGGGATTTGATCTGTCTACTTCCATGCAACGCTTGTGCTTCTGGGTGAGATGCT 411  
QY 370 GTATTTTACCAAGATCTTTTCCGAATGACAGGGGCGAGTTTCAAGCGAGTGTGTGTCTCT 429  
DB 412 GTATTTTACCAAGATCTTTTCCGAATGACAGGGGCGAGTTTCAAGCGAGTGTGTGTCTCT 471  
QY 430 GTCTGCAATTTACAAAGTGTGGGCTTTATCAAGCGCGCCATCAACAAATTTTCCCTGAT 489  
DB 472 GTCTGCGGTTTACGACGCTGCTTGTGCGAGTGTGTGTGCTTCAACGCTTCTCTGAT 531  
QY 490 GATTTCTGGGATTTCCATGCACTGGCTGTGTTTTCAGAGCTGGAAGAGTGTGACAAATGT 549  
DB 532 GACCTGGGAGTTTCTTTCGACTGCTATCGAGAGCTTAAAGAGTGTGACAAATGT 591  
QY 550 GCGTGTGTTTAAAGCTGGGCTCTTCTGCGGCTCTGTTGAGACTCTGCTCTCTCTGCTG 609  
DB 592 GCGTGTGTTTAAAGCTGGGCTCTTCTGCGGCTCTGTTGAGACTCTGCTCTCTCTGCTG 651  
QY 610 GATCAGCGCAAGAGCTTCTGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 669  
DB 652 GATCAGTGAACGAGCTTCTGCGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 711  
QY 670 TGTGTGCAATTTCTCATCTGCTTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 729  
DB 712 CATGTGGCAATCTCTCATCTGCTTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 771  
QY 730 TGTGTGCTCTCAGATATCTGAGCAAGTCTCAGTATCATCAGATTTCTGGCCCGACGAGAA 789  
DB 772 TGTGTGCTCTCAGATATCTGAGCAAGTCTCAGTATCATCAGATTTCTGGCCCGACGAGAA 831  
QY 790 ATGGGCTTTTATGTTGTTGCTCTTATGTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 849  
DB 832 ATGGGCTTTTATGTTGTTGCTCTTATGTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 891  
QY 850 CAAGATCACGTGATGCAAGCGAGTGTGACAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 909  
DB 892 CAAGATCACGTGATGCAAGCGAGTGTGACAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 947  
QY 910 GCTGGGCTTCTGTTGCTAGCAAGATGCTGAGGGGTTGAGGAATTTGGTGTGGTGGG 969  
DB 948 AGTGGGCTTCTGTTGCTAGCAAGATGCTGAGGGGTTGAGGAATTTGGTGTGGTGGG 1007  
QY 970 TGTGTTAAAA 978  
DB 1008 TCTTTTCAA 1016

RESULT 3  
ADN38833  
ID ADN38833 standard; cDNA; 4212 BP.  
XX  
AC ADN38833;  
XX  
DT 17-JUN-2004 (first entry)  
XX  
DE Cancer/angiogenesis/fibrosis-related nucleic acid, SEQ ID NO:151.  
XX  
KW Human; differential expression; cancer; angiogenic disorder;

fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;  
inflammatory disease; autoimmune disease;  
retinal neovascularisation syndrome; scarring; uterine fibroid;  
detection; diagnosis; prognosis; drug screening; drug targeting;  
wound healing; contraception; cytostatic; cardiac; immunomodulatory;  
vulnerability; gene therapy; vaccine; gene; ss.

Homo sapiens.

WO2003042661-A2.

22-MAY-2003.

13-NOV-2002; 2002WO-US036810.

13-NOV-2001; 2001US-0350666P.

21-NOV-2001; 2001US-033464P.

29-NOV-2001; 2001US-0334393P.

03-DEC-2001; 2001US-0335394P.

14-DEC-2001; 2001US-0340376P.

08-JAN-2002; 2002US-0347211P.

10-JAN-2002; 2002US-0347349P.

08-FEB-2002; 2002US-0355250P.

13-FEB-2002; 2002US-0356714P.

20-FEB-2002; 2002US-0359077P.

29-MAR-2002; 2002US-0368809P.

04-APR-2002; 2002US-0370110P.

12-APR-2002; 2002US-0372246P.

05-JUN-2002; 2002US-0386614P.

16-JUL-2002; 2002US-0396839P.

22-JUL-2002; 2002US-0397775P.

22-JUL-2002; 2002US-0397845P.

09-SEP-2002; 2002US-0409450P.

(EOSB-) EOS BIOTECHNOLOGY INC.

Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Heverzi PA;

Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;

WPI; 2003-469649/44.

P-PSDB; ADN38834.

Determining the presence or absence of a pathological cell in a patient,  
useful for diagnosing, prognosing or treating cancer, comprises detecting  
a nucleic acid in a biological sample.

Claim 8; SEQ ID NO 151; 1385pp; English.

The invention relates to nucleic acids and proteins (ADN38683-ADN40064)  
whose expression is upregulated or downregulated in specific cancers or  
other diseases such as angiogenic or fibrotic disorders, and to methods  
of determining the presence or absence of a pathological cell in a  
patient by detecting a nucleic acid at least 80% identical to those of  
the invention or by detecting a polypeptide of the invention. The  
invention also relates to expression vectors and host cells comprising a  
nucleic acid of the invention; antibodies which specifically bind a  
polypeptide of the invention; use of such antibodies for drug targeting;  
and methods of screening for modulators of activity or expression of the  
polypeptides and nucleic acids. The nucleic acids, polypeptides,  
antibodies and methods are useful for diagnosing, prognosing and treating  
cancer and other conditions such as psoriasis, ischaemia, heart disease,  
atherosclerosis, inflammatory diseases, autoimmune diseases, retinal  
neovascularisation syndromes, scarring and uterine fibroids. They may  
also be useful in wound healing and in contraception. The present  
sequence represents a nucleic acid sequence of the invention.

Sequence 4212 BP; 1047 A; 894 C; 932 G; 1339 T; 0 U; 0 Other;

Query Match

Best Local Similarity 17.3%; Score 723.8; DB 11; Length 4212;

Matches 823; Conservative 0; Mismatches 142; Indels 4; Gaps 1;

QY 10 GCTGCTCCGATGCTCCAGAGCGGCCATGGGGCCCGCAGCTGTGGGACCACTGGCGGC 69

|    |      |  |      |
|----|------|--|------|
| Db | 62   | GCTGCTCCAAATGCCCGAGTGGCCATGGCGCCGCACTGGTGAGACAGCTGCAGGC          | 121  |
| QY | 70   | TGGCAGTTCGGAGGTGATTTGGTCCAGGACACTACACTATCTGCTGCTGCAATGCGGA       | 129  |
| Db | 122  | TGGTAGCTCGGAGGTGGACTGGTGGCAGGACAACTACCATCTGCTGCTATCGCCGA         | 181  |
| QY | 130  | GTTCTACAAACAGATCAGCAACGCTCTGTTTCTTTTCACTTTACCTCCCATCTGCACTGTGCTT | 189  |
| Db | 182  | GTTCTACAAACAGATCAGCAATGCTTATTTTCACTTTACCTCCCATCTGCACTGTGCTT      | 241  |
| QY | 190  | GTTCCGCCAGTACGMAAGTGTCTTCAACAGGCGCATCTACTTAATAATGACGCTCCTAGT     | 249  |
| Db | 242  | GTTTCGTAGTATGCAACATGCTTCAACAGTGGCATCTACTTAATCTGACTCTTTTGGT       | 301  |
| QY | 250  | TGTAGTGGGGATTGGATCTGTCTACTTCCATGCAAGCTGAGTTTCTTGGGTGAGATGCT      | 309  |
| Db | 302  | TGTAGTGGGAATGGATCGTCTACTTCCATGCAACCCCTTAGTTCTTGGGTGAGATGCT       | 361  |
| QY | 310  | TGATGAATTCGCAATTCGTGGGTCTGATGTGTCTTGGGCGCATGTGTTTCCCAAGAG        | 369  |
| Db | 362  | TGATGAATTCGCAATTCGTGGGTCTGATGTGTCTTGGGCGCATGTGTTTCCCAAGAG        | 421  |
| QY | 370  | GTATTTACAAAGATCTTTCGGAATGACAGGGCGAGTTCAAGGCGAGTGTGTGTCTCT        | 429  |
| Db | 422  | GTATCTACCAAGATCTTTCGGAATGACCGGGTAGTTCAAGTGGTGGTCACTGTCTCT        | 481  |
| QY | 430  | GTCGCAATTAACAAGTCTTGGCGCTTATCAAGCCCGCATCAACAATATTTCCCTGAT        | 489  |
| Db | 482  | GTCGCGGTACGAGCTGCTGGCATTTGTCAAGCTTCCCATCAACAATCTCTCTGAT          | 541  |
| QY | 490  | GATTCGGGACTTCCATGCACTGCGCTGCTTGTTCAGAGCTGAAGAGGTGCAAACT          | 549  |
| Db | 542  | GACCTGGGAGTTCCTTGCACCTGCTCATCCAGAGCTAAAGAGGTGCAAACT              | 601  |
| QY | 550  | CGGTGTGTTAAGCTGGGCTCTTCTGGGCTCTGGTGGACTCTGGGTCTCTTCTGCTG         | 609  |
| Db | 602  | CGGTGTGTTAAGCTGGGCTCTTCTGGGCTCTGGTGGACTCTGGGTCTCTTCTGCTG         | 661  |
| QY | 610  | GATCAGCAGCAAGCCTTCTGTGAGCTGCTCTCCCTCTTCACTTCCCTTACCTGACTG        | 669  |
| Db | 662  | GATCAGTACCGAGCTTCTCGGAGCTGCTGTGCTTCACTTCCCTTACCTGACTG            | 721  |
| QY | 670  | TGTGTGCAATATTCATCTGCTTGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT         | 729  |
| Db | 722  | CATGTGGCACAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT         | 781  |
| QY | 730  | TGATGCTGCTCAGAGATACCTGAGCAAGTTCAGTCAATCAGATTCCTGGCCAGCGAA        | 789  |
| Db | 782  | TGATGCTGCTCAGAGATTCCTGAGCAAGGCTGCTCATCAAGTTCTGGCCCAATGAA         | 841  |
| QY | 790  | ATGGGCTTTTATGCTGCTTATGCTGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCT        | 849  |
| Db | 842  | ATGGGCTTTTATGCTGCTTATGCTGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCT        | 901  |
| QY | 850  | CAAGATCAGTGTGCAAGGCTGACCAAGCTTCTTACTTACTTCTTATTCGAGTGGC          | 909  |
| Db | 902  | CAAGATCAGTGTGCAAGGCTGACCAAGCTTCTTACTTACTTCTTATTCGAGTGGC          | 957  |
| QY | 910  | GCTGGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT         | 969  |
| Db | 958  | AGTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT         | 1017 |
| QY | 970  | TGTTTAAA 978   |      |
| Db | 1018 | TCITTTCAA 1026   |      |

RESULT 4  
ABS54149  
ID ABS54149 standard; cDNA; 828 BP.  
XX  
AC ABS54149;

|                           |   |
|---------------------------|---|
| XX                        | 25-NOV-2002 (first entry)   |
| DT                        | Human cDNA encoding a liver tumour marker protein, CRG-L1.                |
| XX                        |   |
| DE                        | Human; ss; gene; liver cancer; liver tumour; CRG-L1;                      |
| XX                        | hepatocellular cancer; chromosome 9p.                                     |
| KW                        |   |
| KW                        | Homo sapiens.   |
| XX                        |   |
| OS                        |   |
| XX                        | Key Location/Qualifiers   |
| FH                        | 1..828  |
| FT                        | /tag= a   |
| CDS                       | /product= "Liver tumour marker protein, CRG-L1"                           |
| FT                        |   |
| FT                        |   |
| XX                        | US2002115094-A1.  |
| PN                        |   |
| XX                        |   |
| XX                        | 22-AUG-2002.  |
| PD                        |   |
| XX                        | 14-DEC-2001; 2001US-00017410.   |
| PF                        |   |
| XX                        | 14-DEC-2000; 2000US-0255674P.   |
| PR                        |   |
| XX                        | (FARN/) FARNHAM P J.  |
| PA                        | (GRAV/) GRAVEEL C R.  |
| XX                        |   |
| XX                        | Farnham PJ, Graveel CR;   |
| PI                        |   |
| XX                        | WPI; 2002-706409/76.  |
| DR                        | P-PSDB; ABG32881.   |
| DR                        |   |
| XX                        |   |
| XX                        | Novel polypeptide designated as CRG-L1, useful as diagnostic marker for   |
| PT                        | liver cancer, is differentially expressed in liver tumors relative to     |
| FT                        | normal liver tissues.   |
| FT                        |   |
| XX                        |   |
| PS                        | Claim 2; Page 8; 11pp; English.   |
| XX                        |   |
| CC                        | The invention relates to a polypeptide designated as CRG-L1), which is    |
| CC                        | differentially expressed in liver tumours relative to normal expression   |
| CC                        | in normal liver tissues, designated CRG-L1. Also included are the         |
| CC                        | encoding polynucleotides (in the case of the human sequence, mapping to   |
| CC                        | chromosome 9p), expression constructs, host cells, anti-CRG-L1            |
| CC                        | antibodies, identifying modulators of CRG-L1, and the use of the CRG-L1   |
| CC                        | sequence in the diagnosis of hepatocellular cancer in tumour cells from a |
| CC                        | liver of a human or non-human animal. The CRG-L1 protein and              |
| CC                        | polynucleotide are useful as diagnostic markers for a liver cancer in     |
| CC                        | humans and non-human animals, and as a system for assessing putative      |
| CC                        | therapeutic agents. The present sequence encodes human CRG-L1             |
| XX                        |   |
| SQ                        | Sequence 828 BP; 157 A; 216 C; 211 G; 244 T; 0 U; 0 Other;                |
|                           |   |
| Query Match               | 15.9%; Score 664.8; DB 6; Length 828;                                     |
| Best Local Similarity     | 87.7%; Pred. No. 4.7e-173;  |
| Matches 726; Conservative | 0; Mismatches 102; Indels 0; Gaps 0                                       |
| QY                        | 35 ATGGCGGCCCGCCTGTGGTGGACCACTTCGGGGCTGGCAGTTCCGAGTGGAATTGGTGC 94         |
| Dd                        | 1 ATGGCGGCCCGCCTGTGGTGGACCACTTCGGGGCTGGCAGTTCCGAGTGGAATTGGTGC 60          |
| QY                        | 95 GAGGCAACTACATCATCGTGGCTGCCATTCGGAGTTCTACAACACCATCAGCAACGTC 154         |
| Dd                        | 61 GAGGCAACTACACCATCGTGGCTGCTATTCGCCGAGTTCTACAACACCATCAGCAATGTC 120       |
| QY                        | 155 TTGTTTTTCAVTTTACCCTCCCACATCTGCATGTGCTTGTTCGCCGAGTAGCAACGTCGCTTC 214   |
| Dd                        | 121 TTATTTTTCAVTTTACCOCGCCATCTGCATGTGCTTGTGTGATGATGATGCAACATGCTTG 180     |
| QY                        | 215 AACAGGGCATCTACTTAATATGAGCGCTCCCTAGTTGTAGTGGGATTTGGATCTGTCTAC 274      |
| Dd                        | 181 AACAGTGCATCTACTTAATCTGGACTCTTTTGGTTGTAGTGGGAATTTGGATCCCGTCTAC 240     |
| QY                        | 275 TTCCCATGCAACGCTGAGTTTCTCTGGGTCCAGATGCTTTGATGAACTTGCCATTCGTGGGTT 334   |



XX SQ Sequence 1623 BP; 305 A; 457 C; 382 G; 479 T; 0 U; 0 Other;

Query Match 10.2%; Score 424.2; DB 10; Length 1623;  
Best Local Similarity 81.2%; Pred. No. 3.3e-106;  
Matches 506; Conservative 0; Mismatches 113; Indels 4; Gaps 1;

QY 356 TGGTTTCCAGGAGTATTACCAAGATCTTTCCGAATGACAGGGCAGGTTCAAGGCA 415  
DB 920 TTGTTTCGTCAAGGATCTTACCAAGATCTTTCCGAATGACAGGGGAGGTTCAAGGTG 979

QY 416 GTGGTGTCTCTCTCTGCAATACCAAGTCTTGGCGTTTATCAAGCCCGCATCAAC 475  
DB 980 GTGGTGTCTCTCTCTGCAATACCAAGTCTTGGCGTTTATCAAGCCCGCATCAAC 1039

QY 476 AATATTCCCTGATGATCTTGGGACATTCATGACATGCGTGTGTGTCAGAGCTGAAG 535  
DB 1040 AACATCTCTGATGACCTGGAGTTCTTGGCTGCACTGTCTATCAAGAGCTAAAG 1099

QY 536 AGGTGTGACAAATGTGGTGTGTTAAGCTGGGCTCTTCTCTGGCCTCTGGTGGACCTCG 595  
DB 1100 AGGTGTGACAAATGTGGTGTGTTAAGCTGGGCTCTTCTCTGGCCTCTGGTGGACCTCG 1159

QY 596 GCTCTCTCTGCTGATCAGCAGCAAGCCTTCTGTGAGCTGCTCTCTCTTCACTTC 655  
DB 1160 GCCCTGTCTGCTGATCAGTACCGAGCTTTCTGCGAGCTGTCTCATCTTCACTTC 1219

QY 656 CCTTACCTGCACTGTGTGGCATATTTCTCATCTGCTTGTCTGTAACCTGGGCTGTGTG 715  
DB 1220 CCTTACCTGCACTGTGTGGCATATTTCTCATCTGCTTGTCTGTAACCTGGGCTGTGTG 1279

QY 716 TGCTTCCCTACTTTGATGCTGCTCAGAGATCTGAGCAGAGTCCAGTCACTCAGATTC 775  
DB 1280 TGCTTCCCTACTTTGATGCTGCTCAGAGATCTGAGCAGAGTCCAGTCACTCAGATTC 1339

QY 776 TGGCCAGCAGAGAAATGGCTTTTATGCTGCTTATGCTGCTTCTCTGTGCTCCAC 835  
DB 1340 TGGCCCAATGAGAAATGGCTTTTATGCTGCTTATGCTGCTTCTCTGTGCTCCAC 1399

QY 836 AAGAAGTCCGAGTCAAGATCAGTATGTCGCAAGGATGACAGCTTCTTCTTACTTACTT 895  
DB 1400 AAGAATCATCAGTCAAGATCAGTATGTCGCAAGGATGACAGCTTCTTCTTACTTACTT 1459

QY 896 CTATTGAGTCCGCTGGCTTCTGTTGCTAGCAAGATGCTGAGGGGCTTGAAGAT 955  
DB 1460 CCCCTC---ATGCAGTGGGCTTCTTGTGAGGAAGACGCCAAGGGAGTTCGAATAGT 1515

QY 956 TGGTGTGTGTGGTGTGTTTAA 978  
DB 1516 TGGGTGAGGGCTATGTGTTAA 1538

RESULT 7  
ID ABK62212/c  
XX ABK62212 standard; cDNA; 595 BP.  
AC ABK62212;  
XX  
XX  
XX 18-JUN-2002 (first entry)  
XX  
XX DE Rat sequence differentially expressed in response to a hepatotoxin #119.  
XX  
XX Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;  
KW differential expression; centrilobular necrosis; steatosis.  
XX  
XX OS Rattus norvegicus.  
XX  
XX PN WO200210453-A2.  
XX  
XX PD 07-FEB-2002.  
XX  
XX PF 30-JUL-2001; 2001WO-US023872.  
XX

31-JUL-2000; 2000US-0222040P.  
02-NOV-2000; 2000US-0244880P.  
11-MAY-2001; 2001US-0290029P.  
15-MAY-2001; 2001US-0290645P.  
22-MAY-2001; 2001US-0292336P.  
06-JUN-2001; 2001US-0295798P.  
13-JUN-2001; 2001US-0297457P.  
19-JUN-2001; 2001US-0298884P.  
09-JUL-2001; 2001US-0303459P.  
(GENE-) GENE LOGIC INC.  
Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;  
WPI; 2002-241625/29.  
Predicting toxic effects of compounds or the progression of these toxic effects by determining the changes in gene expression in tissues or cells exposed to the toxin and comparing these to gene expression in unexposed tissues or cells.  
Claim 1; SEQ ID NO 119; 239pp; English.  
The invention relates to methods for predicting toxic effects of compounds or the progression of these toxic effects by determining the global changes in gene expression in tissues or cells exposed to the toxin and comparing these to gene expression in unexposed tissues or cells. Also included are methods of predicting at least one toxic effect of a compound or progression of a toxic effect, preferably the level of hepatotoxicity of a compound, comprising detecting the level of expression in a tissue or cell sample exposed to the compound of two or more genes listed in the specification, where differential expression of the genes is indicative of at least one toxic effect or progression. The method can also be used to identify an agent which modulates the toxic response and predict cellular pathways that a compound modulates in a cell. The methods utilize a set of at least two probes (on a solid support in kit form), where each of the probes comprises a sequence that specifically hybridises to a gene listed in the specification, a computer system comprising a database containing information identifying the expression level in a tissue or cell sample exposed to a hepatotoxin of a set of genes comprising at least two genes listed in the specification, and a user interface to view the information used to present information identifying the expression level in a tissue or cell of at least one gene listed in the specification. The method is useful for elucidating global changes in gene expression and for identifying toxicity markers in tissues or cell exposed to a known toxin. The genes may be used as toxicity markers in drug screening and toxicity assays. The genes and gene expression information may be used as diagnostic markers for the prediction or identification of the physiological state of tissue or cell sample that has been exposed to a compound or agent. Hepatotoxicity is characterised by centrilobular necrosis and steatosis. The present sequence is an expressed sequence tag (EST) or cDNA derived from a gene which is differentially expressed in response to a hepatotoxic agent  
SQ Sequence 595 BP; 197 A; 101 C; 146 G; 151 T; 0 U; 0 Other;  
Query Match 9.6%; Score 400.4; DB 6; Length 595;  
Best Local Similarity 87.9%; Pred. No. 7.2e-100;  
Matches 531; Conservative 0; Mismatches 56; Indels 17; Gaps 8;  
QY 3557 ATTCTATTGGTTGGATTTCTCAGGAGTCTGAAATCTTCTCCACGAGTCTCTTCCCAA 3616  
DB 595 ATTCTATTGGTTGGATTTCTCAGGAGTCTGAAATCTTCTCCACGAGTCTCTTCCCAA 536  
QY 3617 CCCCTACCCAGAGCAACACTGGGATTTGAACCTTC---TCTGACTCTTCTTCTCCCTC 3672  
DB 535 CCGCTCCAGATC---CACTGGGATTTGAACCTTCAAATCTGCTCTCTTCTTCTTC 479  
QY 3673 AGGTCTGACAACTAATGGTCTCTGGGGAACCCAGCTAGGGCTTCCCAACTCTT-AT 3731  
DB 478 AGGTCTGACAACTAATGGTCTCTGGGGAACCCAGCTAGGGCTTCCCAACTCTTCTT 419  
QY 3732 CCAGCTGAACCTGGATTTCTTCCCAACAGGGCTTGACCTGGGTGCTGTGGTCCCACTGG 3791



|    |   |  |  |
|----|---|--|--|
| DT | 06-MAY-2004   | (first entry)                          |  |
| XX | Human 3T3 cell conversion promoter PP11646 cDNA.  |  |  |
| XX | 3T3 cell conversion; promoter; human; gene; ss.   |  |  |
| KW | Homo sapiens.   |  |  |
| OS | CN1403477-A.  |  |  |
| XX | 19-MAR-2003.  |  |  |
| XX | 12-SEP-2001; 2001CN-00126725.   |  |  |
| PF | 12-SEP-2001; 2001CN-00126725.   |  |  |
| XX | (SHAN-) SHANGHAI XINSHIJI GENE TECHN DEV CO LTD.  |  |  |
| XX | Gu J, Yang S;   |  |  |
| XX | WPI; 2003-494226/47.  |  |  |
| DR | P-PSDB; ADL06641.   |  |  |
| XX | Human protein with function of promoting 3T3 cell conversion and its coding sequence.   |  |  |
| PT | Example 1; SEQ ID NO 1; 41pp; Chinese.  |  |  |
| XX | This invention describes a novel human protein with 3T3 cell conversion promoting function, polynucleotides encoding the polypeptide and the recombinant process of producing the polypeptide. The present invention also discloses the agonist resisting the polypeptide and its treatment effect. The present invention also discloses the application of the polynucleotides encoding the human protein with 3T3 cell conversion promoting function. |  |  |
| CC | Sequence 1527 BP; 376 A; 357 C; 383 G; 411 T; 0 U; 0 Other;   |  |  |
| XX | Query Match   | 7.8%; Score 323.6; DB 10; Length 1527; |  |
| CC | Best Local Similarity   | 73.5%; Pred. No. 2.4e-78;              |  |
| CC | Matches 475; Conservative   | 0; Mismatches 74; Indels 97; Gaps 1;   |  |
| QY | 142 GATCAGCAACGCTCTTTTTCATTTTACCTCCCATCTGCATGCTGTTCCGCCAGTA   | 201                                    |  |
| Db | 178 GATCAGCAATGCTTTTTCATTTTACCTCCCATCTGCATGCTGTTCCGCCAGTA   | 237                                    |  |
| QY | 202 CGCAACGCTGTTTCAACAGCGGCATCTACTTAATATGACGCTCTTGTAGTGGGAT   | 261                                    |  |
| Db | 238 TGCAACATGCTTCAACAGTGGCATCTACTTAATCTGGACTCTTTTGGTTGTAGTGGGAT   | 297                                    |  |
| QY | 262 TGGATCTGCTACTTCCATGCAACGCTGATTTCTGGTTCAGATGCTTGATGAATTGC  | 321                                    |  |
| Db | 298 TGGATCCGCTACTTCCATGCAACCTTAGTTTCTGGTTCAGATGCTTGATGAATTGC  | 357                                    |  |
| QY | 322 CATCTGNGGGTTCGATGTGCTTTGGCCATGTTTCCAGAGAGTATTTACCAAA  | 381                                    |  |
| Db | 358 AGTCCTTTGGGTTCTGATGTGCTTTGGCCATGTTTCCCAAGAGTATTTACCAAA  | 417                                    |  |
| QY | 382 GATCTTTCCGAATGACAGGGCAGGTTCAAGCAGTGTGTGTCCTGTCTGCAATATAC  | 441                                    |  |
| Db | 418 GATCTTTCCGAATGACCGGGTAGGTTCAAGGTGGTGGTCACTGCTCTGCTGCGGTAC   | 477                                    |  |
| QY | 442 AACGTGCTGGGTTTATCAAGCCGCCATCAACATATTTCCCTGATGATTTCTGGACT  | 501                                    |  |
| Db | 478 GACGTGCTGGGATTTGTCAAGCCGCCATCAACATCTCTCTGATGACCTGGGAGT  | 537                                    |  |
| QY | 502 TCCATGCACTGGCTGCTTTGTCAGAGCTGAAG-----   | 535                                    |  |
| Db | 538 TCCATGCACTGCTGCTCTCATCGAGCTTAAGAGGATGAGAGCAACCGAGGAG  | 597                                    |  |
| QY | 536 -----   | 535                                    |  |
| Db | 598 ACACAGGAAGGTGGCCAGAGGTTGGAGACAAAGGTTCTGACGATGATGACTCTCTG  | 657                                    |  |

|    |     |   |     |
|----|-----|---|-----|
| QY | 536 | -----AGGTGACAAATGTCGTGTTTAAAGCTGGGCTCTTCTCTGGCCCTCT       | 584 |
| Db | 658 | GACCCCGTGCAGGTGACAAACATGCTGTTTAAAGCTGGGCTCTTCTCTGGCCCTCT  | 717 |
| QY | 585 | GGTGGACTCTGGCTCTCTTCTGCTGGATCAGCGACCAAGCCTTCTGTAGCTGTCTCT | 644 |
| Db | 718 | GGTGGACCTGGCCCTCTTCTGCTGGATCAGTGACCGAGCTTCTGCGAGCTGTCTCAT | 777 |
| QY | 645 | CTTTTCACTTCCCTACCTGACATGCTGTGGGATATTTCTCATCTG             | 690 |
| Db | 778 | CTTTCAACTTCCCTACCTGACATGCTGTGGGATATTTCTCATCTG             | 823 |

RESULT 10

|          |   |  |
|----------|---|--|
| ADL06642 | ADL06642  | standard; DNA; 1527 BP.                |
| XX       | ADL06642;   |  |
| XX       | 06-MAY-2004   | (first entry)                          |
| XX       | Human 3T3 cell conversion promoter PP11646 DNA.   |  |
| XX       | 3T3 cell conversion; promoter; human; gene; ds.   |  |
| XX       | Homo sapiens.   |  |
| XX       | Key   | Location/Qualifiers                    |
| FT       | CDS   | 218..640                               |
| FT       |   | /*tag= a                               |
| FT       |   | /product= "PP11646"                    |
| XX       | CN1403477-A.  |  |
| XX       | 19-MAR-2003.  |  |
| XX       | 12-SEP-2001; 2001CN-00126725.   |  |
| XX       | 12-SEP-2001; 2001CN-00126725.   |  |
| XX       | (SHAN-) SHANGHAI XINSHIJI GENE TECHN DEV CO LTD.  |  |
| XX       | Gu J, Yang S;   |  |
| XX       | WPI; 2003-494226/47.  |  |
| DR       | P-PSDB; ADL06641.   |  |
| XX       | Human protein with function of promoting 3T3 cell conversion and its coding sequence.   |  |
| PT       | Example 1; SEQ ID NO 3; 41pp; Chinese.  |  |
| XX       | This invention describes a novel human protein with 3T3 cell conversion promoting function, polynucleotides encoding the polypeptide and the recombinant process of producing the polypeptide. The present invention also discloses the agonist resisting the polypeptide and its treatment effect. The present invention also discloses the application of the polynucleotides encoding the human protein with 3T3 cell conversion promoting function. |  |
| CC       | Sequence 1527 BP; 376 A; 357 C; 383 G; 411 T; 0 U; 0 Other;   |  |
| XX       | Query Match   | 7.8%; Score 323.6; DB 10; Length 1527; |
| CC       | Best Local Similarity   | 73.5%; Pred. No. 2.4e-78;              |
| CC       | Matches 475; Conservative   | 0; Mismatches 74; Indels 97; Gaps 1;   |
| QY       | 142 GATCAGCAACGCTCTTTTTCATTTTACCTCCCATCTGCATGCTGTTCCGCCAGTA   | 201                                    |
| Db       | 178 GATCAGCAATGCTTTTTCATTTTACCTCCCATCTGCATGCTGTTCCGCCAGTA   | 237                                    |
| QY       | 202 CGCAACGCTGTTTCAACAGCGGCATCTACTTAATATGACGCTCTTGTAGTGGGAT   | 261                                    |
| Db       | 238 TGCAACATGCTTCAACAGTGGCATCTACTTAATCTGGACTCTTTTGGTTGTAGTGGGAT   | 297                                    |
| QY       | 262 TGGATCTGCTACTTCCATGCAACGCTGATTTCTGGTTCAGATGCTTGATGAATTGC  | 321                                    |
| Db       | 298 TGGATCCGCTACTTCCATGCAACCTTAGTTTCTGGTTCAGATGCTTGATGAATTGC  | 357                                    |
| QY       | 322 CATCTGNGGGTTCGATGTGCTTTGGCCATGTTTCCAGAGAGTATTTACCAAA  | 381                                    |
| Db       | 358 AGTCCTTTGGGTTCTGATGTGCTTTGGCCATGTTTCCCAAGAGTATTTACCAAA  | 417                                    |
| QY       | 382 GATCTTTCCGAATGACAGGGCAGGTTCAAGCAGTGTGTGTCCTGTCTGCAATATAC  | 441                                    |
| Db       | 418 GATCTTTCCGAATGACCGGGTAGGTTCAAGGTGGTGGTCACTGCTCTGCTGCGGTAC   | 477                                    |
| QY       | 442 AACGTGCTGGGTTTATCAAGCCGCCATCAACATATTTCCCTGATGATTTCTGGACT  | 501                                    |
| Db       | 478 GACGTGCTGGGATTTGTCAAGCCGCCATCAACATCTCTCTGATGACCTGGGAGT  | 537                                    |
| QY       | 502 TCCATGCACTGGCTGCTTTGTCAGAGCTGAAG-----   | 535                                    |
| Db       | 538 TCCATGCACTGCTGCTCTCATCGAGCTTAAGAGGATGAGAGCAACCGAGGAG  | 597                                    |
| QY       | 536 -----   | 535                                    |
| Db       | 598 ACACAGGAAGGTGGCCAGAGGTTGGAGACAAAGGTTCTGACGATGATGACTCTCTG  | 657                                    |



Db 238 TGCAACATGCTTCAACAGTGGCATCTACTTAATCTGGACTCTTTGGTTGTAGTGGGAAT 297  
Qy 262 TGGATCTGCTACTTCCATGCAAGCTGAGTTTCTGGGTGATGCTTGTATGAATTCG 321  
Db 298 TGGATCCGCTACTTCCATGCAACCTTAGTTTCTGGGTGATGCTTGTATGAATTCG 357  
Qy 322 CATTCCTGCTGGGTTCCTGATGTGCTTTGGCCATGCTTTCCAGAGGTATTTACCAA 381  
Db 358 AGTCCTTTGGGTTCCTGATGTGCTTTGGCCATGCTTTCCAGAGGTATTTACCAA 417  
Qy 382 GATCTTTCGGAATGACAGGGCAGTTCCTCAAGCAGTGTGTGCTCTCTGCAATATAC 441  
Db 418 GATCTTTCGGAATGACAGGGGTAGGTTCAGGTGTGTGCTCTCTGCGGTAC 477  
Qy 442 AACGTGCTGGCTTATCAAGCCGCCATCAACAATTTCCCTGATGATTTCTGGGACT 501  
Db 478 GACGTGCTGGCTTATCAAGCCGCCATCAACAATTTCTCTGATGACCTGGGAGT 537  
Qy 502 TCCATGCACTGGCTGCTTGTTCAGAGCTGAAG----- 535  
Db 538 TCCTTGCACTGCACTGCTCATCGCAGAGTAAAGAGGCATGAGAGAACCGAAGGAG 597  
Qy 536 ----- 535  
Db 598 ACACAGAAAGGTGCCAGCAGGAGGTGGAGACAAGGTCTGACGATGAGTCTCTG 657  
Qy 536 -----AGGTGACATGCGGTGTTTAACTGGGCTCTTCTGCGCTCT 584  
Db 558 GACCCCGTGGAGGTGACAACTGCGTGTGTTTAACTGGGCTCTTCTGCGCTCT 717  
Qy 585 GGTGACTCTGGCTCTCTTCTGCTGATCAGCGACCAAGCCTTCTGAGCTGCTCTCT 644  
Db 718 GGTGACCTGGCTCTTCTGCTGATCAGTACCGAGCTTTCTGCGAGCTGCTGTAT 777  
Qy 645 CCTTTCACTCCCTACTGCACTGTGTGTGGCATATTTCTCATCTG 690  
Db 778 CTTTCAACTTCCCTACTGCACTGATGTGTGGAGCGTTCAGATG 823

RESULT 11

AD121468  
ID AD121468 standard; cDNA; 487 BP.  
XX AC AD121468;  
XX AC AD121468;  
XX DT 15-APR-2004 (first entry)  
XX DE Novel human expressed sequence tag, EST #167.  
XX KW forensic; nutritional source; damaged tissue; diseased tissue;  
KW myeloid cell disorder; lymphoid cell disorder;  
KW bone cartilage tissue growth; tendon tissue growth;  
KW ligament tissue growth; nerve tissue growth; regeneration; wound healing;  
KW tissue repair; tissue replacement; burn; incision; ulcer; cancer; human;  
KW ss; gene; EST; expressed sequence tag.  
XX OS Homo sapiens.  
XX PN WO2003025148-A2.  
XX PD 27-MAR-2003.  
XX PF 19-SEP-2002; 2002WO-US029964.  
XX PR 19-SEP-2001; 2001US-0323739P.  
XX PR 13-SEP-2002; 2002US-00323739.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
PI Ghosh M, Xue Au, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang D;  
PI Haley-Vicente D;  
XX

DR WPI; 2003-354603/33.  
DR P-PSDB; ADI21688.  
XX  
PT New polynucleotides and secreted proteins, useful for treating myeloid or  
PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve  
PT tissue growth or regeneration, in wound healing, and in tissue repair and  
PT replacement.  
XX  
PS Example 2; SEQ ID NO 719; 156bp; English.  
XX  
CC The invention relates to an isolated polynucleotide encoding a  
CC polypeptide with biological activity. The polynucleotides and  
CC polypeptides are useful in diagnostics, forensics, gene mapping,  
CC identification of mutations responsible for genetic disorders and other  
CC traits, to assess biodiversity, as nutritional sources or supplements.  
CC The polynucleotides may also be used as molecular weight markers,  
CC chromosome markers or map related gene positions, or as an antigen to  
CC raise anti-DNA antibodies or elicit immune response. The polypeptides are  
CC useful for raising antibodies, as markers for tissues in which the  
CC corresponding polypeptide is expressed, for re-engineering damaged or  
CC diseased tissues, for treating myeloid or lymphoid cell disorders, in  
CC bone cartilage, tendon, ligament and/or nerve tissue growth or  
CC regeneration, in wound healing, in tissue repair and replacement, in  
CC healing of burns, incisions and ulcers, and in treating cancer. The  
CC present sequence represents a novel human expressed sequence tag, EST.  
XX  
SQ Sequence 487 BP; 81 A; 138 C; 131 G; 134 T; 0 U; 3 Other;

Query Match 7.1%; Score 294.8; DB 10; Length 487;  
Best Local Similarity 89.5%; Pred. NO. 1.2e-70;  
Matches 317; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 10 GCTGCTCCGATGCTCCAGAGCGGCATGGGCGCCCGCACCTGTGGGACCACTGCGGC 69  
Db 134 GCTGCTCCGATGCTCCAGAGCGGCATGGGCGCCCGCACCTGTGGGACCACTGCGGC 193  
Qy 70 TGGCAGTTTCGAGTGGATTTGGTTCGAGGACCACTACACTATCGTCCCTGCAATTCGCGA 129  
Db 194 TGGTAGCTCGGAGTGGACTGGTTCGAGGACCACTACACTATCGTCCCTGCAATTCGCGA 253  
Qy 130 GTTCTCAACACAGATCAGCAGCTTGTGTTTTCATTTTACCTCCCATCTGCGATGCTGTT 189  
Db 254 GTTCTCAACACAGATCAGCAGCTTGTGTTTTCATTTTACCTCCCATCTGCGATGCTGTT 313  
Qy 190 GTTCCGCGAGTACGCAACAGTGTCTTCAACAGCGCATCTACTTAATATGGAGCTCCTAGT 249  
Db 314 GTTTCGTGATGCAACATGCTTCAACAGTGGCATCTACTTAATCTGGACTCTTTTGGT 373  
Qy 250 TGTAGTGGGATTTGGATCTGTCTACTTCCATGCAACGCTGAGTTTCCTGGGTGAGATGCT 309  
Db 374 TGTAGTGGGAAATTGGATCCGCTACTTCCATGCAACCTTAGTTTCTGGGTGAGATGCT 433  
Qy 310 TGATGAACTTGGCAATTCCTGTTGGTTCCTGATGTGCTTTGGCCATGTTGTTTC 363  
Db 434 TGATGAACTTGGCAATTCCTGTTGGTTCCTGATGTGCTTTAGCCATGTTGTTTCCC 487

RESULT 12

AAK51566  
ID AAK51566 standard; cDNA; 1215 BP.  
XX AC AAK51566;  
XX AC AAK51566;  
XX DT 06-NOV-2001 (first entry)  
XX DE Human polynucleotide SEQ ID NO 111.  
XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation; ss.  
XX  
OS Homo sapiens.



PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman I, Goodrich R;  
XX WPI; 2001-476283/51.  
DR P-PSDB; AAM/9417.  
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful  
PT in diagnosis and gene therapy.  
XX Claim 1; Page 4462; 6221pp; English.  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication  
XX  
SQ Sequence 1631 BP; 247 A; 399 C; 344 G; 350 T; 0 U; 291 Other;

Query Match 4.2%; Score 176.8; DB 4; Length 1631;  
Best Local Similarity 51.3%; Pred. No. 1.2e-37;  
Matches 437; Conservative 0; Mismatches 412; Indels 3; Gaps 1;

QY 74 AGTTCGAGGTGATGCTGGGAGGACAACTACACTATGTCGCTGCCATTGCGGATTC 133  
Db 80 AGTTCGAGGTGATGCTGGGAGGACAACTACACTATGTCGCTGCCATTGCGGATTC 139  
QY 134 TACAACACGATCAGACGCTCTGTTTTCATTTTACCTCCATCTGCATGCTGCTTC 193  
Db 140 TACAACACGTTTCCATATATCCCTCTTTCATCTTCGGGCCACTGATGCTCCTGATG 199  
QY 194 CGCCAGTACGCAAGCTGCTTCAACAGCGGATCTACTTAATATGAGCGCTCTAGTTGTA 253  
Db 200 CACCCGATGCCAGAGCGCTCCGCTACATTTACGTTGCTGGGCTCTTTCATGATC 259  
QY 254 GTGGGATTCGATGCTCTACTTCCATGCAACGCTGATTTCTGGGTCAGATGCTTGTAT 313  
Db 260 ATAGGCTGTTTCCATGATTTCCACATGACGCTCAGCTTCTTGGGCCAGCTGCTGGAC 319  
QY 314 GAACCTTCCCATCTGTCGGTCTGATGCTGCTTGGCCATGTTGTTTCCAGGAGGTAT 373  
Db 320 GAGATGCCATCTGTGGCTCTGGGCGATGGCTATAGCATATGATGCCCGCTGCTAT 379  
QY 374 TTACCAAGATCTTTTGGAAATGACAGGGGAGGTTCAAGGCGAGTGTGTGCTCTGCT 433  
Db 380 TTCCCTCTCTCTTCTGGGGGAAACAGGTCCTCAGTTTCTATCCGCTGTTCTTCAACCACT 439  
QY 434 GCAATTACACGCTGCTGGGTTTATCAAGCCGCCATCAACATATTTCCCTGATGATT 493  
Db 440 GTGGTCAGCACCTTCTGCTCTTCTTCCGCGCCCAAGCTCAACGCTTACGCTTCAACAGC 499  
QY 494 CTGGGACTTCCATGCACTGCGCTGCTTGTGTCAGAGCTGAAGAGGTGTGACAAATGTCGT 553  
Db 500 ATTGCCCTGCACATCTCTACATCGTGTGCCAGGATGACAGGACCAATTAAGAG 559  
QY 554 GTGTTTAAGCTGGGCTCTTCTCTGGGCTCTGGTGGACTCTGGCTCTCTTCTGCTGGATC 613  
Db 560 CTTCGGCACCTGATGAGGCTCTCGTGGTTTATGGGCTGTGCTCTGACCACTGATC 619  
QY 614 AGGACCAACCTTCTGTAGGCTCTCTCTCTTCACTTCCCTACCTGCACTGTGTG 673  
Db 620 AGTGACCGTCTGCTTGTGACGCTTCTGGCAGAGGATTCATTTCTTCTATGTCACAGCATC 679  
QY 674 TGGCATATCTCATCTGCTCTGCTTCTGATCCTGGGCTGTGCTGCTGCTGCTTGTAT 733

Db 680 TGGCATGTGCTCATCAGCATCACCTTCCCTTATGGCATGGTCAACATGGCCCTTGGTGGAT 739  
QY 734 GCTGCTTCAGAGATACCTGAGCAAGTCCAGTTCATCAGATTCTGGCCAGCGAGAAATGG 793  
Db 740 GCCAACTATGAGATGCCAGGTGAACCTCAAAGTCCGCTACTGCTCTGGGACAGTTGG 799  
QY 794 GCTTTTATGTTGCTCCCTTATGTTGCTCTTCTGTTGCCCAAGAAAGTCCGCCAGTCAAG 853  
Db 800 CC---CGTGGGCTGCTCTACGTGGAATCGGGGTGATGACAAAGGACTGCTGAGACCTG 856  
QY 854 ATCAGGTGATGGCAAGCAGTGCACGCTTCTCTACTTACTTCTTATTCGAGTGGCGCTG 913  
Db 857 CCAGCTCTTGACTATCCACCACCCCAACAACTTGGCTGTGTTGAGAAAGATAGCCCCG 916  
QY 914 GGCTTCGTTTTCG 925  
Db 917 GTCAGGACTTGC 928

RESULT 14  
AAH48639  
ID AAH48639 standard; DNA; 792 BP.  
XX  
AC AAH48639;  
XX  
DT 21-SEP-2001 (first entry)  
XX  
DE Human ceramidase K1 DNA.  
XX  
KW Ceramidase; human; K1; antiproliferative; anticancer; anti-eczema;  
KW antiporiasis; dermatological; ceramide; sphingosine; treatment; ds;  
KW altered cell proliferation; cancer; neurodermatitis; eczema; cosmetic;  
KW permeability barrier; psoriasis; ichthyosis; lamellar ichthyosis ICR2.  
XX  
OS Homo sapiens.  
XX  
PN WO200155408-A1.  
XX  
PD 02-AUG-2001.  
XX  
PF 27-JAN-2001; 2001WO-EP000900.  
XX  
PR 27-JAN-2000; 2000DE-01003293.  
PR 09-MAR-2000; 2000DE-01011392.  
XX  
PA (MEMO-) MEMOREC MEDICAL MOLECULAR RES COLOGNE ST.  
XX  
PI Hofmann K, Conradt M;  
XX  
XX WPI; 2001-483256/52.  
DR P-PSDB; AAB86363.  
XX  
XX New ceramidase containing specific structural motif, useful for  
PT diagnosis, prevention and treatment of ceramidase defects, e.g.  
PT ichthyosis, also in cosmetics.  
XX  
XX Claim 3; Page 16; 31pp; German.  
CC  
CC This invention describes novel human ceramidase (I) containing a specific  
CC structural motif which has antiproliferative, anticancer, anti-eczema,  
CC antiporiasis and dermatological activity. (I) cleave ceramide, resulting  
CC in formation of sphingosine. (I), or the nucleic acid (II) encoding it,  
CC are used for diagnosis, prevention or treatment of diseases associated  
CC with ceramide defects, particularly altered cell proliferation (cancer)  
CC or altered ceramide layers on the skin (neurodermatitis, eczema,  
CC psoriasis), also for targeted modification of the permeability barrier by  
CC ceramidase or its activators, e.g. for transcutaneous delivery of  
CC substances. Specifically they are used for diagnosis of ichthyosis,  
CC particularly lamellar ichthyosis ICR2. (I) can also be used in cosmetics.  
CC This sequence encodes human ceramidase K1  
XX  
SQ Sequence 792 BP; 145 A; 241 C; 200 G; 206 T; 0 U; 0 Other;

Query Match 4.2%; Score 174.8; DB 4; Length 792;  
Best Local Similarity 52.6%; Pred. No. 2.8e-37; Indels 0; Gaps 0;  
Matches 380; Conservative 0; Mismatches 342;

QY 74 AGTTCGGAGTGGATGGTGGAGGACAACTACATATCGTGCCTGCCATTCGCCAGTTTC 133  
DB 25 AGCTCCGAGTGGAGTGGTGGAGGACAACTTCAGTACTCGGAGCTGGTGGCCGAGTTTC 84

QY 134 TACACACAGATCAGCAAGCTCTGTTTTCATTTTACCTCCCATCTGCATGTGCTGTTTC 193  
DB 85 TACAACACAGCTTCTCCAATATCCCTCTTCTATCTTCGGGCCACTGATGATGCTCTCTGATG 144

QY 194 CGCCAGTACGCAAGCTCTTCAACAGCGGCATCTACTTAATATGAGCGCTCTAGTTGA 253  
DB 145 CACCCGATGCCCCAGAGCGCTCCCGCTACATTTAGTTGTTCTGTTGCTCTCTTCATGATC 204

QY 254 GTGGGATTTGATCTGTCTACTTCCATGCAAGCGCTGAGTTTCTCGGTCAGATGCTTTGAT 313  
DB 205 ATAGGCTGTTCTCCATGTAATTTCCACATGACGCTCAGCTTCTCGGCCAGCTGCTGGAC 264

QY 314 GAATTCGCCATCTGTGGTCTGATGTGTGCTTTGGCCATGTGTTTCCAGAGGTAT 373  
DB 265 GAGATGCGCATCTGTGGCTCTCGGCACTGAGTATGATATGATGATGATGATGATGATGATG 324

QY 374 TTACCAAGATCTTTTCGGAATGACAGGGGAGGTTTCAAGGCGAGTGTGTGCTGCTCT 433  
DB 325 TTCCCTCTCTTCTTGGGGGAGAGGTCAGGTTCCAGTTTATCGGCTGCTCTTCATCACCAC 384

QY 434 GCAATTACAACTGCTTTGGCGTTTATCAAGCCCGCCATCAACAATATTTCCCTGATGAT 493  
DB 385 GTGGTCAGCACCTTCTGTCTTCTGCGGCCACAGGTCACGCTTACGCTTCAACAGC 444

QY 494 CTGGAGCTTCATGCACTGCTGCTGTGTGAGAGCTGAAGAGGTGTGACATGTGCT 553  
DB 445 ATTGCCCTGCACATCTCTACATCGTGTGCGGAGTACAGGAGGACCAATATAGGAG 504

QY 554 GTGTTTAACTGGGCTCTTCTGCGCTCTGTTGAGCTCTGCTCTCTCTCTGCTGATC 613  
DB 505 CTTCGACCTGATGAGGCTCTCGGTTTATGAGGCTGTGCTCTGACAGCTGATC 564

QY 614 AGCGACCAAGCTTCTGTGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 673  
DB 565 AGTGACCGTCTGTTTGCAGCTTCTGGCAGAGGATTCATTTCTCTATCTGCACAGCATC 624

QY 674 TGGCATTTCTCATCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 733  
DB 625 TGGCATGTGCTCATCAGCATCCTTCTCTTATGCGATGCTCACCATGCGCTTGGTGGAT 684

QY 734 GCTGCTCAGATACCTCAGCAAGGTCAGATCTCAGATCTCAGATCTCAGATCTCAGATCT 793  
DB 685 GCCAATATGATGCGAGGTGAAACCTTCAAGTCCGCTACTGGCTCGGACAGTTGG 744

QY 794 GC 795  
DB 745 CC 746

RESULT 15  
ABL24043  
ID ABL24043 standard; DNA; 537 BP.  
XX ABL24043;  
XX  
XX  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 23602.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.

XX 27-SEP-2001.  
PD  
XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR  
XX 11-JUL-2000; 2000US-00614150.  
PR  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 23602; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB2072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 537 BP; 102 A; 146 C; 148 G; 141 T; 0 U; 0 Other;  
SQ

Query Match 3.8%; Score 160.6; DB 4; Length 537;  
Best Local Similarity 58.1%; Pred. No. 1.9e-33;  
Matches 283; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

QY 53 TGGGACCACTCGCGGCTGGAGTTTGGAGGTTGGATTGTGGAGGACAACTACACTATC 112  
DB 49 TGGGAGCACTTAAGACCCGGAAGTCTGCCCTGCTGACTGGTGGAGGAACTACTTGATT 108

QY 113 GTGCCCTGCAATTCGCCAGTTCTACACAGCATGACGACGCTTGTGTTTTCATTTTACCT 172  
DB 109 TCGTCCAACTCGCGGAGTTCGTGAACAGCTTTAGCACTTCTGTTCTACTCTGCGC 168

QY 173 CCAATCTGATGCTGTTGTTCCGCACTAGCAACGCTTTCACAGCGCATCTACTTA 232  
DB 169 CCGCTCTTATATGCTCTTCAAGGAGTACGACGCTTTGTGACGCCCGGAATCCAGTC 228

QY 233 ATATGACGCTCTCTAGTTGTAGTGGGATTTGGATCTGTCTACTTCCATGCAACGCTGAGT 292  
DB 229 ATCTGGGTGCTGTCATCTGTTGGCTGAGTTCGATGATCTTCCATGCCACTTTGAGT 288

QY 293 TTCCTGGGTGATGATGCACTTGCATTCGCTGCTGGGTTCGTGATGTTGTTGGCC 352  
DB 289 CTGATTGGCCAGCTGCTGGACGAACTGGCCATCTGCGTCTTCTATGCGGCTTTTCG 348

QY 353 ATGTGTTTCCGAGGAGTATTTACCAAGATCTTTCGGAATTCACAGGGGCGAGTTCAAG 412  
DB 349 CTCCTTCTATCCGAAGCGATATCTCCAAAGTTCGTAAGAAACGATCGCAAAACCTTCAGT 408

QY 413 GCAGTGGTGTGCTCTGCTGCTGCAATTCACAACTGCTGCGGTTTATCAAGCCCGCATC 472  
DB 409 TGGCTCATGCTCTTCTGCGGATTCGTCGACGGCTTCTGTTGGTGAAGCCATTGTT 468

QY 473 AACAAATTTTCCCTGATGATTTTGGGACTTTCGAGACTTCCATGCACTGCGCTGCTTGTGAGAGCTG 532  
DB 469 AACGCTTTTGTCTCATGTTTCATGATGTTGCGGACCATGTTATGTTCTTACACAGAGCTG 528

QY 533 AAGAGCT 539  
DB 529 CAGAGGT 535

Search completed: November 20, 2004, 21:28:44  
Job time : 1777.81 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: November 20, 2004, 20:51:27 ; Search time 357.166 Seconds  
(without alignments)  
8308.591 Million cell updates/sec  
Title: US-10-017-410-1  
Perfect score: 4175  
Sequence: 1 ggcacagggtctcgat.....aaaaaaaaaaaaaaaaaac 4175  
Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 824507 seqs, 355394441 residues  
Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES  |       |             |        |        |    |                     |                     |  |  |  |
|------------|-------|-------------|--------|--------|----|---------------------|---------------------|--|--|--|
| Result No. | Score | Query Match | Length | DB     | ID | Description         |                     |  |  |  |
|            |       |             |        |        |    |                     |                     |  |  |  |
| C          | 1     | 114.2       | 2.7    | 358    | 4  | US-09-621-976-11241 | Sequence 11241, A   |  |  |  |
|            | 2     | 55.8        | 1.3    | 7218   | 1  | US-08-232-463-14    | Sequence 14, Appl   |  |  |  |
|            | 3     | 46          | 1.1    | 1141   | 4  | US-09-806-708B-22   | Sequence 22, Appl   |  |  |  |
|            | 4     | 45.4        | 1.1    | 1063   | 4  | US-09-799-451-875   | Sequence 875, Appl  |  |  |  |
| C          | 5     | 44.8        | 1.1    | 832    | 4  | US-09-621-976-2813  | Sequence 2813, Appl |  |  |  |
|            | 6     | 42.4        | 1.0    | 832    | 4  | US-09-621-976-2813  | Sequence 2813, Appl |  |  |  |
|            | 7     | 41.8        | 1.0    | 30310  | 4  | US-09-657-346A-96   | Sequence 96, Appl   |  |  |  |
|            | 8     | 41          | 1.0    | 289    | 3  | US-09-007-005-17    | Sequence 17, Appl   |  |  |  |
| C          | 9     | 41          | 1.0    | 289    | 3  | US-09-244-796-17    | Sequence 17, Appl   |  |  |  |
|            | 10    | 40.8        | 1.0    | 435    | 4  | US-09-248-796A-6940 | Sequence 6940, Appl |  |  |  |
|            | 11    | 40          | 1.0    | 1044   | 4  | US-09-495-406-12    | Sequence 12, Appl   |  |  |  |
|            | 12    | 40          | 1.0    | 1044   | 4  | US-09-816-028A-16   | Sequence 16, Appl   |  |  |  |
| C          | 13    | 40          | 1.0    | 1044   | 4  | US-10-303-162-16    | Sequence 16, Appl   |  |  |  |
|            | 14    | 40          | 1.0    | 11474  | 4  | US-09-495-406-1     | Sequence 1, Appli   |  |  |  |
|            | 15    | 40          | 1.0    | 11474  | 4  | US-09-816-028A-1    | Sequence 1, Appli   |  |  |  |
|            | 16    | 40          | 1.0    | 11474  | 4  | US-10-303-162-1     | Sequence 1, Appli   |  |  |  |
| C          | 17    | 39.2        | 0.9    | 1141   | 4  | US-09-806-708B-22   | Sequence 22, Appl   |  |  |  |
|            | 18    | 38.8        | 0.9    | 302    | 3  | US-09-328-111-679   | Sequence 679, Appl  |  |  |  |
|            | 19    | 38.4        | 0.9    | 2044   | 4  | US-09-327-138C-38   | Sequence 38, Appl   |  |  |  |
|            | 20    | 38.4        | 0.9    | 2863   | 4  | US-09-327-138C-2    | Sequence 2, Appli   |  |  |  |
| C          | 21    | 38.4        | 0.9    | 3646   | 4  | US-09-327-138C-36   | Sequence 36, Appl   |  |  |  |
|            | 22    | 38.4        | 0.9    | 3665   | 4  | US-09-327-138C-35   | Sequence 35, Appl   |  |  |  |
|            | 23    | 38.4        | 0.9    | 3686   | 4  | US-09-327-138C-3    | Sequence 3, Appli   |  |  |  |
|            | 24    | 38.4        | 0.9    | 3686   | 4  | US-09-327-138C-37   | Sequence 37, Appl   |  |  |  |
| C          | 25    | 38.4        | 0.9    | 3940   | 4  | US-09-327-138C-1    | Sequence 1, Appli   |  |  |  |
|            | 26    | 38.2        | 0.9    | 168174 | 4  | US-10-071-411A-63   | Sequence 63, Appl   |  |  |  |
|            | 27    | 38.2        | 0.9    | 168273 | 4  | US-10-071-411A-2    | Sequence 2, Appli   |  |  |  |

Sequence 90, Appl  
Sequence 17, Appl  
Sequence 126, App  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 3, Appli  
Sequence 10, Appl  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 9223, Ap  
Sequence 45, Appl  
Sequence 45, Appl  
Sequence 29557, A  
Sequence 13559, A  
Sequence 157, App  
Sequence 157, App  
Sequence 157, App

ALIGNMENTS

RESULT 1  
US-09-621-976-11241/c  
; Sequence 11241, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Giordano, J.Y.  
; FILE OF INVENTION: ESTS and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621.976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.Pm  
; SEQ ID NO 11241  
; LENGTH: 358  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-11241

Query Match 2.7%; Score 114.2; DB 4; Length 358;  
Best Local Similarity 90.4%; Pred. No. 2.1e-25;  
Matches 122; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
Qy 10 GCTCTCCGATGCTCCAGAGCGCGCCCGCCGCGCTGTGGGACACCTGCGGGC 69  
Db 136 GCTGCTCAATGCCCCGAGTGGCCATGGGCGCCCGCGCTGTGGGACACCTGCGGGC 77  
Qy 70 TGGCAGTTCCGAGGTGGATTGGTCGCGAGGACAACTACACTATCGTGCCTGCGCATTCGCGA 129  
Db 76 TGGTAGCTCGGAGTGGACTGGTTCGAGGACAACTACACTATCGTGCCTGCGCATTCGCGA 17  
Qy 130 GTTCTACACACGAT 144  
Db 16 GTTCTACACACGAT 2

RESULT 2  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500









```
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match          1.0%; Score 41; DB 3; Length 289;
Best Local Similarity 7.6%; Pred. No. 0.026;
Matches 17; Conservative 97; Mismatches 111; Indels 0; Gaps 0;

QY 541 TGCAATGTGCGTGTAAAGCTGGGCTCTTCTCTGGCCCTCTGGTGACTCTGGCTCT 600
Db 233 YGCVAYGCTGTGYSYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNY 174
QY 601 CTTCTGCTGGATCAGCGACCAAGCCTTCTGTGAGCTGCTCTCTCTTCACTCCCTA 660
Db 173 YNYSYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNY 114
QY 661 CTTGACCTGTGTGGCATAATCTCATCTGCCTCTGCTGCTGGCTGGCTGTGCTT 720
Db 113 YNYSYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNY 54
QY 721 GCGCTACTTGTGCTGCTCAGATACCTGAGCAAGTCCAGT 765
Db 53 YTYTGTYAYATYTYGTYAYAYAYTYAYTYAYTYTYGY 9

RESULT 10
US-09-248-796A-6940/c
; Sequence 6940, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 6940
; LENGTH: 435
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (387),(391)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unkno
US-09-248-796A-6940

Query Match          1.0%; Score 40.8; DB 4; Length 435;
Best Local Similarity 47.3%; Pred. No. 0.042;
Matches 123; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 2251 TAACCTGCTGGCTATTTCTTGTATATAGTCACAAATATTTATAGAAATGAAAGTGTA 2310
Db 330 TGAATAGCTGTGTGTGTGTATTTATTTATTTATTTATTTATTTATTTATTTATTT 271
QY 2311 AATAAGTTGCTTTCTTTATTAATAATTCATAGCATTTTCACTCTCCAGGTAGCAACAC 2370
```

```
Db 270 AATATCTTGATTAATTTAATTTGATTAATTTTCAAAATCTATAGTAATTAAGTCAA 211
QY 2371 AGTCATTTTGATGGCGATTACAGGCGAGGAATCTTGCTTTTGAATGATTTGATATGTC 2430
Db 210 TTCCAATAATTTACGCAATAAAGATTTCATCTAGATCAATATTTCTCTTTATTTTAA 151
QY 2431 ATTCTACTGTAGCTTAAATAATGTTTAAATGACCTTTGAGTGAAATGTTTGGCAACCT 2490
Db 150 ATATTGTTGATAAATTTACCAAAATGTAATAATTTCTAATGATTTAAATAATTTGTAGTAGAAG 91
QY 2491 AGGGCTTTATGATCAGAAAT 2510
Db 90 AATGGAAGGATAAGAGAAAT 71

RESULT 11
US-09-495-406-12/c
; Sequence 12, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; FILE REFERENCE: 019633-0001100S
; CURRENT APPLICATION NUMBER: US/09/495,406
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 1044
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1044)
; OTHER INFORMATION: beta-1,4 N-acetylgalactosaminyl (GalNAC)
; OTHER INFORMATION: transferase from C. jejuni strain OH4384 (ORF 5a
; OTHER INFORMATION: of LOS biosynthesis locus)
US-09-495-406-12

Query Match          1.0%; Score 40; DB 4; Length 1044;
Best Local Similarity 45.7%; Pred. No. 0.16;
Matches 139; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 3862 AAAACATGGCTTGCTCTTTCCATACACATCGGTCAGATTAAACTATTTTATAACCA 3921
Db 963 ATAGAAATTTTAAATTTCTTCAATACTTATCCAAATTAATTTATTAATAATATCTTGAGCTCT 904
QY 3922 CAGGAATTTAAACCAAGCAATAGAGTACTTTTCAGATATAAACTGTGTTTCATCTTTATG 3981
Db 903 ATGTTTTTTTAAAGGAAATGANAATTAATTTAATTCAGTGTGAATAATATCTCTGTG 844
QY 3982 TAGAGTGTGCTATGATAGGCGGTATGTACCTCGCTGAAGTAATTAATTAACATAGCTCT 4041
Db 843 TTTATACTTTAAATGCTCTAAAGAGTGTATATTTGAAGAAAACTTTGTTGTTGCTTT 784
QY 4042 GGGAGGATTTACAGACCTTTTGCACTTATGCTTTTTTTTGGAACCTGTGATAACCATGTC 4101
Db 783 TTTCCAATTTGTTATTTGATTTCTGTCAATTAATACTTTCTTCCATTTTAAGTTATAATTAT 724
QY 4102 AATATTAAGCAATAACTGGCATTTTCTGTGAATAAACAATGATATGATCTATAAAAAA 4161
Db 723 GATAAGCAAGCAATCTCTGTTCTTTTCAATATATATGCAATTAATTTATTTTACAAAGAA 664
QY 4162 AAAA 4165
Db 663 AAAA 660
```

## RESULT 12

US-09-816-028A-16/c  
; Sequence 16, Application US/09816028A  
; Patent No. 6699705  
; GENERAL INFORMATION:  
; APPLICANT: Gilbert, Michel  
; APPLICANT: Wakarchuk, Warren W.  
; APPLICANT: National Research Council of Canada  
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of  
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics  
; FILE REFERENCE: 019633-000111US  
; CURRENT APPLICATION NUMBER: US/09/816,028A  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/118,213  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: US 09/495,406  
; PRIOR FILING DATE: 2000-01-31  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 1044  
; TYPE: DNA  
; ORGANISM: Campylobacter jejuni  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1044)  
; OTHER INFORMATION: Beta-1,4 N-acetylgalactosaminyl (GalNAc)  
; OTHER INFORMATION: transferase from C. jejuni strain OH4384 (ORF 5a)  
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)  
US-09-816-028A-16

Query Match 1.0%; Score 40; DB 4; Length 1044;

Best Local Similarity 45.7%; Pred. No. 0.16; Indels 0; Gaps 0;  
Matches 139; Conservative 0; Mismatches 165;  
QY 3862 AAAACATGGCTGTGTCTTTCCATACACATCCGGTCAGATTAAACCTATTTTATAACCA 3921  
Db 963 ATAGAAATTTTAAATTTCTTCAATCTATCCAAATATATTTATAAATATCTTGAGCTCT 904  
QY 3922 CAGAAATTAACCAAGCAATAGAGTACTTTCAGATATAAAGTGTGTTTCACTTTATG 3981  
Db 903 ATGTTTTTTTTAAAAAGGAAATGATAATTTATTTAATTCAGTGTGAATAATATCCCTGTG 844  
QY 3982 TAGAGTGTGCTATGTATAGGGGGTATGACCTCGCTGAAGTAAATATTAACCATAGCTCT 4041  
Db 843 TTTATCTTTAATTTGCTCTAAGAGTGATATTTGAAGAAATACTTTGTTTGTGCTTT 784  
QY 4042 GGGAGGATTTACAGACCTTTTTCACCTTTTATGCTTTTGTGAACCTCTGATAACCATGGTC 4101  
Db 783 TTTCCAAATGTTATTTGATTTCTGCAATTAATTAATTTTCCATTTTAAGTTATAATTATT 724  
QY 4102 AATATTAAGCAATAAGTGCATTTCTGTGAATAAATGATATGATATCTATAAAAAA 4161  
Db 723 GATAAGCAAGCAATCTCTGTTCTTTCAATATATATGTCATTATTTTACAAAGAAA 664  
QY 4162 AAAA 4165  
Db 663 AAAA 660

## RESULT 13

US-10-303-162-16/c  
; Sequence 16, Application US/10303162  
; Patent No. 6723545  
; GENERAL INFORMATION:  
; APPLICANT: Gilbert, Michel  
; APPLICANT: Wakarchuk, Warren W.  
; APPLICANT: National Research Council of Canada  
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of  
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics  
; FILE REFERENCE: 019633-000111US  
; CURRENT APPLICATION NUMBER: US/10/303,162  
; CURRENT FILING DATE: 2002-11-21

; PRIOR APPLICATION NUMBER: US/09/816,028  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/118,213  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: US 09/495,406  
; PRIOR FILING DATE: 2000-01-31  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 1044  
; TYPE: DNA  
; ORGANISM: Campylobacter jejuni  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1044)  
; OTHER INFORMATION: Beta-1,4 N-acetylgalactosaminyl (GalNAc)  
; OTHER INFORMATION: transferase from C. jejuni strain OH4384 (ORF 5a)  
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)  
US-10-303-162-16

Query Match 1.0%; Score 40; DB 4; Length 1044;

Best Local Similarity 45.7%; Pred. No. 0.16; Indels 0; Gaps 0;  
Matches 139; Conservative 0; Mismatches 165;  
QY 3862 AAAACATGGCTGTGTCTTTCCATACACATCCGGTCAGATTAAACCTATTTTATAACCA 3921  
Db 963 ATAGAAATTTTAAATTTCTTCAATCTATCCAAATATATTTATAAATATCTTGAGCTCT 904  
QY 3922 CAGAAATTAACCAAGCAATAGAGTACTTTCAGATATAAAGTGTGTTTCACTTTATG 3981  
Db 903 ATGTTTTTTTTAAAAAGGAAATGATAATTTATTTAATTCAGTGTGAATAATATCCCTGTG 844  
QY 3982 TAGAGTGTGCTATGTATAGGGGGTATGACCTCGCTGAAGTAAATATTAACCATAGCTCT 4041  
Db 843 TTTATCTTTAATTTGCTCTAAGAGTGATATTTGAAGAAATACTTTGTTTGTGCTTT 784  
QY 4042 GGGAGGATTTACAGACCTTTTTCACCTTTTATGCTTTTGTGAACCTCTGATAACCATGGTC 4101  
Db 783 TTTCCAAATGTTATTTGATTTCTGCAATTAATTAATTTTCCATTTTAAGTTATAATTATT 724  
QY 4102 AATATTAAGCAATAAGTGCATTTCTGTGAATAAATGATATGATATCTATAAAAAA 4161  
Db 723 GATAAGCAAGCAATCTCTGTTCTTTCAATATATATGTCATTATTTTACAAAGAAA 664  
QY 4162 AAAA 4165  
Db 663 AAAA 660

## RESULT 14

US-09-495-406-1/c  
; Sequence 1, Application US/09495406  
; Patent No. 6503744  
; GENERAL INFORMATION:  
; APPLICANT: Wakarchuk, Warren W.  
; APPLICANT: National Research Council of Canada  
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of  
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics  
; FILE REFERENCE: 019633-000111US  
; CURRENT APPLICATION NUMBER: US/09/495,406  
; CURRENT FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: US 60/118,213  
; PRIOR FILING DATE: 1999-02-01  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 11474  
; TYPE: DNA  
; ORGANISM: Campylobacter jejuni  
; FEATURE:  
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384  
; OTHER INFORMATION: including LOS biosynthesis locus



**This Page Blank (uspto)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: November 20, 2004, 20:53:12 ; Search time 1950.22 Seconds  
(without alignments)  
11567.962 Million cell updates/sec

Title: US-10-017-410-1  
Perfect score: 4175  
Sequence: 1 ggcacgagggtgtccgat.....aaaaaaaaaaaaaaaaaaac 4175  
Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 3627888 seqs, 2701811610 residues  
Total number of hits satisfying chosen parameters: 7255776  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCT05\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
20: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
21: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description          |
|------------|-------|-------------|--------|-------|----------------------|
| 1          | 4175  | 100.0       | 4175   | 13    | US-10-017-410-1      |
| 2          | 723.8 | 17.3        | 4202   | 16    | US-10-188-832-21     |
| 3          | 723.8 | 17.3        | 4212   | 15    | US-10-295-027-151    |
| 4          | 664.8 | 15.9        | 828    | 13    | US-10-017-410-3      |
| 5          | 653.8 | 15.7        | 822    | 15    | US-10-182-447-6      |
| 6          | 400.4 | 9.6         | 595    | 9     | US-09-917-800A-119   |
| 7          | 329.2 | 7.9         | 744    | 10    | US-09-945-527-48     |
| 8          | 174.8 | 4.2         | 792    | 15    | US-10-182-447-4      |
| 9          | 131.2 | 3.1         | 1098   | 15    | US-10-172-118-2618   |
| 10         | 131.2 | 3.1         | 1098   | 16    | US-10-342-887-2618   |
| 11         | 125.8 | 3.0         | 1473   | 15    | US-10-106-698-611    |
| 12         | 65    | 1.6         | 418    | 16    | US-10-242-535A-50197 |

|    |      |     |        |    |                      |                   |
|----|------|-----|--------|----|----------------------|-------------------|
| 13 | 65   | 1.6 | 418    | 16 | US-10-085-783A-50197 | Sequence 50197, A |
| 14 | 63.6 | 1.5 | 77530  | 17 | US-10-322-281-265    | Sequence 265, App |
| 15 | 61.4 | 1.5 | 659158 | 9  | US-09-771-208-20     | Sequence 20, Appl |
| 16 | 58.2 | 1.4 | 35425  | 15 | US-10-017-161-2429   | Sequence 2429, Ap |
| 17 | 58.2 | 1.4 | 35425  | 15 | US-10-292-798-2069   | Sequence 2069, Ap |
| 18 | 55.4 | 1.3 | 91609  | 17 | US-10-322-696-109    | Sequence 109, App |
| 19 | 52.6 | 1.3 | 19619  | 15 | US-10-369-214-129    | Sequence 129, App |
| 20 | 49.4 | 1.2 | 7516   | 15 | US-10-311-455-34     | Sequence 34, Appl |
| 21 | 49.4 | 1.2 | 102374 | 13 | US-10-087-192-667    | Sequence 667, App |
| 22 | 48.6 | 1.2 | 50295  | 11 | US-09-997-722-259    | Sequence 259, App |
| 23 | 48.6 | 1.2 | 115780 | 17 | US-10-367-094-96     | Sequence 96, Appl |
| 24 | 48.4 | 1.2 | 60968  | 15 | US-10-085-117-181    | Sequence 181, App |
| 25 | 47.6 | 1.1 | 52479  | 11 | US-09-997-722-61     | Sequence 61, Appl |
| 26 | 47.6 | 1.1 | 97415  | 17 | US-10-322-281-287    | Sequence 287, App |
| 27 | 47.4 | 1.1 | 50657  | 17 | US-10-322-281-601    | Sequence 601, App |
| 28 | 47.4 | 1.1 | 215980 | 10 | US-09-972-546-16     | Sequence 16, Appl |
| 29 | 47   | 1.1 | 32495  | 17 | US-10-322-281-401    | Sequence 401, App |
| 30 | 47   | 1.1 | 67191  | 11 | US-09-997-722-169    | Sequence 169, App |
| 31 | 47   | 1.1 | 67191  | 15 | US-10-105-612-1      | Sequence 1, Appl  |
| 32 | 46.8 | 1.1 | 185555 | 13 | US-10-087-192-1999   | Sequence 1999, Ap |
| 33 | 46.6 | 1.1 | 171936 | 15 | US-10-265-071-24     | Sequence 24, Appl |
| 34 | 46.6 | 1.1 | 171936 | 15 | US-10-025-966A-24    | Sequence 24, Appl |
| 35 | 46.4 | 1.1 | 73507  | 13 | US-10-087-192-1147   | Sequence 1147, Ap |
| 36 | 46   | 1.1 | 96593  | 11 | US-09-997-722-151    | Sequence 151, App |
| 37 | 45.8 | 1.1 | 197775 | 13 | US-10-087-192-853    | Sequence 853, App |
| 38 | 45.6 | 1.1 | 38342  | 16 | US-10-221-714A-472   | Sequence 472, App |
| 39 | 45.6 | 1.1 | 54648  | 15 | US-10-085-117-289    | Sequence 289, App |
| 40 | 45.4 | 1.1 | 801    | 15 | US-10-182-447-5      | Sequence 5, Appl  |
| 41 | 45.4 | 1.1 | 1063   | 16 | US-10-302-172-875    | Sequence 875, App |
| 42 | 45.2 | 1.1 | 10279  | 15 | US-10-311-455-1563   | Sequence 1563, Ap |
| 43 | 45.2 | 1.1 | 10279  | 16 | US-10-240-589C-85    | Sequence 85, Appl |
| 44 | 45.2 | 1.1 | 55906  | 17 | US-10-322-281-417    | Sequence 417, App |
| 45 | 45   | 1.1 | 305    | 9  | US-09-967-769A-63    | Sequence 63, Appl |

ALIGNMENTS

RESULT 1  
US-10-017-410-1  
; Sequence 1, Application US/10017410  
; Publication No. US20020115094A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnham, Peggy J  
; APPLICANT: Graveel, Carrie R  
; TITLE OF INVENTION: Polynucleotide Differentially Expressed in Liver Cancer  
; FILE REFERENCE: 960296.97401  
; CURRENT APPLICATION NUMBER: US/10/017,410  
; CURRENT FILING DATE: 2001-12-14  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4175  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (35)..(859)  
US-10-017-410-1

|                       |                 |   |           |              |
|-----------------------|-----------------|---|-----------|--------------|
| Query Match           | 100.0%;         | Score 4175;   | DB 13;    | Length 4175; |
| Best Local Similarity | 100.0%;         | Pred. No. 0;  |           |              |
| Matches 4175;         | Conservative 0; | Mismatches 0;   | Indels 0; | Gaps 0;      |
| Qy                    | 1               | GGCAGGAGGCTGCTCCGATGCTCCAGAGCGGCCCATGGCGCCGCGACTGGTGGGACCA 60 |           |              |
| Db                    | 1               | GGCAGGAGGCTGCTCCGATGCTCCAGAGCGGCCCATGGCGCCGCGACTGGTGGGACCA 60 |           |              |
| Qy                    | 61              | CCTGCGGGGTGCGAGTTCGGAGTGGTGGCGAGGACAACTACATCTATCGTGCCTGC 120  |           |              |
| Db                    | 61              | CCTGCGGGGTGCGAGTTCGGAGTGGTGGCGAGGACAACTACATCTATCGTGCCTGC 120  |           |              |
| Qy                    | 121             | CATTGCGAGTCTTACACACGATCAGCAACGCTCTTTGTTTTTACCTCCCATCTG 180    |           |              |

Db 121 CATGGCGAGTTCTTAAACAGATCAGCAAGCTCTGTTTTCATTTTACCTCCCACTG 180  
Qy 181 CATGTGTTTTCGCGCAGTACGCAAGTGTCTTCAACAGCGGCATCTACTTAAATATGAC 240  
Db 181 CATGTGTTTTCGCGCAGTACGCAAGTGTCTTCAACAGCGGCATCTACTTAAATATGAC 240  
Qy 241 GCTCCTAGTTAGTGGGATTTGATCTGTCTACTTCCATGCAAGCTGAGTTTCTCTGG 300  
Db 241 GCTCCTAGTTAGTGGGATTTGATCTGTCTACTTCCATGCAAGCTGAGTTTCTCTGG 300  
Qy 301 TCAGATGCTTATGAATTCGATCTGTGGGTTCTGATGTGTCTTGGCCATGTGGT 360  
Db 301 TCAGATGCTTATGAATTCGATCTGTGGGTTCTGATGTGTCTTGGCCATGTGGT 360  
Qy 361 TCCAGAGAGTATTTACAAAGATCTTTTCGAAATGACAGGGGAGGTTCAAGGAGTGGT 420  
Db 361 TCCAGAGAGTATTTACAAAGATCTTTTCGAAATGACAGGGGAGGTTCAAGGAGTGGT 420  
Qy 421 GTGTGCTCTGCTCAATTACAAAGTCTTGGCGTTTATCAAGCCCGCATCAACATAT 480  
Db 421 GTGTGCTCTGCTCAATTACAAAGTCTTGGCGTTTATCAAGCCCGCATCAACATAT 480  
Qy 481 TTCCCTGATGATTTCTGGGACTTCCATGCACTGGCTGCTTGTGAGAGCTGAAGAGGT 540  
Db 481 TTCCCTGATGATTTCTGGGACTTCCATGCACTGGCTGCTTGTGAGAGCTGAAGAGGT 540  
Qy 541 TGACAAATGTGGTGTGTTAAGCTGGGCTCTTCTGCGCTCTGTGTGGAATCTGGCTCT 600  
Db 541 TGACAAATGTGGTGTGTTAAGCTGGGCTCTTCTGCGCTCTGTGTGGAATCTGGCTCT 600  
Qy 601 CTTCTGCTGATCAGGACCAAGCTTCTGTGAGCTGCTCTCCCTTTTCACTTCCCTA 660  
Db 601 CTTCTGCTGATCAGGACCAAGCTTCTGTGAGCTGCTCTCCCTTTTCACTTCCCTA 660  
Qy 661 CTTGCACTGTGTGGCATATTCATCTGCTTCTGCTGCTGCTGCTGCTGCTGCT 720  
Db 661 CTTGCACTGTGTGGCATATTCATCTGCTTCTGCTGCTGCTGCTGCTGCTGCT 720  
Qy 721 CGCCTACTTTGATGCTGCTCAGAGATACCTGAGCAAGTCCAGTCAATCAGATTCGGCC 780  
Db 721 CGCCTACTTTGATGCTGCTCAGAGATACCTGAGCAAGTCCAGTCAATCAGATTCGGCC 780  
Qy 781 CAGCGAGAAATGGGCTTTATTTGTTGCTTCCCTTATGTGCTTCTGCTGCTCAGAGAA 840  
Db 781 CAGCGAGAAATGGGCTTTATTTGTTGCTTCCCTTATGTGCTTCTGCTGCTCAGAGAA 840  
Qy 841 GTCGCCAGTCAAGATCAGTGCATGAGCAAGGAGTCCAGTCAATCAGATTCGGCT 900  
Db 841 GTCGCCAGTCAAGATCAGTGCATGAGCAAGGAGTCCAGTCAATCAGATTCGGCT 900  
Qy 901 CGAGTGGCGCTGGGCTTCTGCTAGCAAGATGGCTGAGGGGTTGAGAAATGGTG 960  
Db 901 CGAGTGGCGCTGGGCTTCTGCTAGCAAGATGGCTGAGGGGTTGAGAAATGGTG 960  
Qy 961 TGGTGGGCTTTTAAATCTGCTTCTGCTGATCTAACTGGAACCACTGCTGCTGCT 1020  
Db 961 TGGTGGGCTTTTAAATCTGCTTCTGCTGATCTAACTGGAACCACTGCTGCTGCT 1020  
Qy 1021 CCCTAGGTTAAGTAGAGGCTCAGGACATAAGTGTCTTCCCTCAGTATCTTCCAGGC 1080  
Db 1021 CCCTAGGTTAAGTAGAGGCTCAGGACATAAGTGTCTTCCCTCAGTATCTTCCAGGC 1080  
Qy 1081 ATAGGGCTGCTGGGTTATCTCCATATGATCATCAAGAGTAGTCTTTGGGAGACT 1140  
Db 1081 ATAGGGCTGCTGGGTTATCTCCATATGATCATCAAGAGTAGTCTTTGGGAGACT 1140  
Qy 1141 AGGCAACCCCAAGTTCTTGTGAGAGGCTAGCTTGTCTTCCACATTTCCAGACT 1200  
Db 1141 AGGCAACCCCAAGTTCTTGTGAGAGGCTAGCTTGTCTTCCACATTTCCAGACT 1200  
Qy 1201 CTCACCCCTGATAAACAAGCTTCTGATTTGGTATGAAGGTTTCCAGAACTTTTCA 1260

Db 1201 CTCACCCCTGATAAACAAGCTTCTGATTTGGTATGAAGGTTCCAGAACTTTTCA 1260  
Qy 1261 TTTTGGCGGAGAAACTGTCTTCAACAAACCAAGTGGGCAAAACACGTTGGGGGTG 1320  
Db 1261 TTTTGGCGGAGAAACTGTCTTCAACAAACCAAGTGGGCAAAACACGTTGGGGGTG 1320  
Qy 1321 TGACTAAGACGGGTGGCTTGTCACTCGCACCTTCTAGCTTTCCAGAGTCTTCCAGCTGT 1380  
Db 1321 TGACTAAGACGGGTGGCTTGTCACTCGCACCTTCTAGCTTTCCAGAGTCTTCCAGCTGT 1380  
Qy 1381 TGTGGATCTTCCACTACTCTGCTGAGAGATGAGCCACCGCTTCCAGAGGCAAGCTG 1440  
Db 1381 TGTGGATCTTCCACTACTCTGCTGAGAGATGAGCCACCGCTTCCAGAGGCAAGCTG 1440  
Qy 1441 GCAACACCTCTATGCCAAAGCTACACTCTCTTTAAGCACACATTTACATAGACATAT 1500  
Db 1441 GCAACACCTCTATGCCAAAGCTACACTCTCTTTAAGCACACATTTACATAGACATAT 1500  
Qy 1501 TTCTGCTCTTCCAGAGTGCAGACGCTCAGACCCACAGAGAACTCCTTCAGGTATGTGAA 1560  
Db 1501 TTCTGCTCTTCCAGAGTGCAGACGCTCAGACCCACAGAGAACTCCTTCAGGTATGTGAA 1560  
Qy 1561 GATTCCACACACAGCCCCCTCTTGTGACCTCTGTGAGAGCAAGGCTCTGTTGTAATCGCG 1620  
Db 1561 GATTCCACACACAGCCCCCTCTTGTGACCTCTGTGAGAGCAAGGCTCTGTTGTAATCGCG 1620  
Qy 1621 GAGGAGCTCTGAGAGCTTGAATCTGAGAGGCTTGAAGATTTTGAAGCTTTTCTTA 1680  
Db 1621 GAGGAGCTCTGAGAGCTTGAATCTGAGAGGCTTGAAGATTTTGAAGCTTTTCTTA 1680  
Qy 1681 TTGGAAGCTTTTGT 1740  
Db 1681 TTGGAAGCTTTTGT 1740  
Qy 1741 AGGGATATGTGCTTCTGAGTGTAGCTTCCCTAGGATGTGGGCTCCGGCTTTTGT 1800  
Db 1741 AGGGATATGTGCTTCTGAGTGTAGCTTCCCTAGGATGTGGGCTCCGGCTTTTGT 1800  
Qy 1801 TTTTCAATACCTGGTGTAACTGTTTCTCAGAGCACTTGTCTTCTTCTGATGCTGGC 1860  
Db 1801 TTTTCAATACCTGGTGTAACTGTTTCTCAGAGCACTTGTCTTCTTCTGATGCTGGC 1860  
Qy 1861 GGTCACTACACTCTGATTTGGCTGGGATCCAGGGAGAAAGGGGAGACTCTGACA 1920  
Db 1861 GGTCACTACACTCTGATTTGGCTGGGATCCAGGGAGAAAGGGGAGACTCTGACA 1920  
Qy 1921 TGAATGTCTCTCACCTGCATCTACTGTCTTCTCAGTGGGCTCTCTTCAAGGATGAAGTGG 1980  
Db 1921 TGAATGTCTCTCACCTGCATCTACTGTCTTCTCAGTGGGCTCTCTTCAAGGATGAAGTGG 1980  
Qy 1981 GTATGGGGTATCTCAGGGATGTTTGAACCTCAGGACCTTCTGCTTCTGACATTTCCAT 2040  
Db 1981 GTATGGGGTATCTCAGGGATGTTTGAACCTCAGGACCTTCTGCTTCTGACATTTCCAT 2040  
Qy 2041 GTCAGTGTGAGACTGCACTCAGGGGACTGAGTTTGAACACCCGTTCTGAGTGGTCTCT 2100  
Db 2041 GTCAGTGTGAGACTGCACTCAGGGGACTGAGTTTGAACACCCGTTCTGAGTGGTCTCT 2100  
Qy 2101 GAGTCTGAAGGAGTGTAGTGGGTTCTGGCACTTCTAGGATCTCTTACTCTGTTTGAAG 2160  
Db 2101 GAGTCTGAAGGAGTGTAGTGGGTTCTGGCACTTCTAGGATCTCTTACTCTGTTTGAAG 2160  
Qy 2161 CCTTCAAGGGTACAAGAGTGGAACTGGACCTTTAAAGAGTTTAAATGAATAGACTTCAT 2220  
Db 2161 CCTTCAAGGGTACAAGAGTGGAACTGGACCTTTAAAGAGTTTAAATGAATAGACTTCAT 2220  
Qy 2221 TCTGCTTTTGTGGCTTTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACT 2280  
Db 2221 TCTGCTTTTGTGGCTTTTGAAGTAACTTGAAGTAACTTGAAGTAACTTGAAGTAACT 2280  
Qy 2281 CACAAATATTTATAGACATGAAGGTTAAATGAAGTGTCTTTTATTAATTCATAGCA 2340  
Db 2281 CACAAATATTTATAGACATGAAGGTTAAATGAAGTGTCTTTTATTAATTCATAGCA 2340



QY 2341 TTTACCAACTCTCCAGGTAGCAAAACACACAGTCATTGTGATGCGGATTTTCAAGGCGAG 2400  
DB 2341 TTTACCAACTCTCCAGGTAGCAAAACACACAGTCATTGTGATGCGGATTTTCAAGGCGAG 2400  
QY 2401 AATACCTGGTTTGAATGATGTATATGTCTATTTTACCTGAGCTTAAATATGTTAAAA 2460  
DB 2401 AATACCTGGTTTGAATGATGTATATGTCTATTTTACCTGAGCTTAAATATGTTAAAA 2460  
QY 2461 TGACTTTGAGTGAATGTTCTGGCAACTAGGGGTTTATGGATCAGAAATGCGGCTGAG 2520  
DB 2461 TGACTTTGAGTGAATGTTCTGGCAACTAGGGGTTTATGGATCAGAAATGCGGCTGAG 2520  
QY 2521 TCCTTTGGTTTTGAGTCTAGAAGTTTTCAGAGGCAAAATCAAACTAGCCTGTGTTCTG 2580  
DB 2521 TCCTTTGGTTTTGAGTCTAGAAGTTTTCAGAGGCAAAATCAAACTAGCCTGTGTTCTG 2580  
QY 2581 GTTCATCTTACAGGCTCGGAGCTGCCCTTATATATTTCTACATAAGGACTTATATACA 2640  
DB 2581 GTTCATCTTACAGGCTCGGAGCTGCCCTTATATATTTCTACATAAGGACTTATATACA 2640  
QY 2641 TAAGTCTGTATAAATGCTCTGAAGATGACACTAGCTGCCCTTCACTCGAAGGGTCTGCT 2700  
DB 2641 TAAGTCTGTATAAATGCTCTGAAGATGACACTAGCTGCCCTTCACTCGAAGGGTCTGCT 2700  
QY 2701 GGGCTCGAGAGTTGGTTCAACAGTTAAGAACCCGCTGTGCTCTCCAGACGACCCAG 2760  
DB 2701 GGGCTCGAGAGTTGGTTCAACAGTTAAGAACCCGCTGTGCTCTCCAGACGACCCAG 2760  
QY 2761 CGACTTGTAGGCTCATAAACAGCAACTCCAGCCCTGGGGCATCTGACGCCCTCTTCTGA 2820  
DB 2761 CGACTTGTAGGCTCATAAACAGCAACTCCAGCCCTGGGGCATCTGACGCCCTCTTCTGA 2820  
QY 2821 CTTCTAAGGGCATTTGGTGACACGCTACATACAGGCGCAAAACATTTATACATGTAAAC 2880  
DB 2821 CTTCTAAGGGCATTTGGTGACACGCTACATACAGGCGCAAAACATTTATACATGTAAAC 2880  
QY 2881 GTAATAAATGATAGTTAGTGAGACGGCTGAAGAAAGGAGTTTATAGTCAAGGTTTA 2940  
DB 2881 GTAATAAATGATAGTTAGTGAGACGGCTGAAGAAAGGAGTTTATAGTCAAGGTTTA 2940  
QY 2941 GTCTGACCTGAGTGCCTTCTTAGCCGGATGTTCTCCTCAAGTCAGGGAAGAACCAT 3000  
DB 2941 GTCTGACCTGAGTGCCTTCTTAGCCGGATGTTCTCCTCAAGTCAGGGAAGAACCAT 3000  
QY 3001 TTCTTTTATCAGGAAAGGACTTATTCCTTAGGGCCCTCTGCTGACATCCCTAGGAACA 3060  
DB 3001 TTCTTTTATCAGGAAAGGACTTATTCCTTAGGGCCCTCTGCTGACATCCCTAGGAACA 3060  
QY 3061 GAGATAAATACGGATGGATGTAATGAACATGCTTGGGTGAAGAGCGGAGTACCTGA 3120  
DB 3061 GAGATAAATACGGATGGATGTAATGAACATGCTTGGGTGAAGAGCGGAGTACCTGA 3120  
QY 3121 CTGGACCCAGTGGGCCACTTCCACAGCGAAGCCCGGTAACCGATGTCACCTCCAGAAATC 3180  
DB 3121 CTGGACCCAGTGGGCCACTTCCACAGCGAAGCCCGGTAACCGATGTCACCTCCAGAAATC 3180  
QY 3181 TTCTCCCTTCTGGTAGAGATGACATCGATGATGCTGATGATGCTGACCACTGGGCCCTGCAATG 3240  
DB 3181 TTCTCCCTTCTGGTAGAGATGACATCGATGATGCTGATGATGCTGACCACTGGGCCCTGCAATG 3240  
QY 3241 GGTGTCGGGCTGTGTTCTGTTCCGAATCTACCTGAGATCTAGGACAGAGGAGCCATGA 3300  
DB 3241 GGTGTCGGGCTGTGTTCTGTTCCGAATCTACCTGAGATCTAGGACAGAGGAGCCATGA 3300  
QY 3301 ATGTTACCAAGTGTATGCTGCTGATGTTTACAGTTTGAAGTATGTTGTTT 3360  
DB 3301 ATGTTACCAAGTGTATGCTGCTGATGTTTGAAGTATGTTGTTT 3360  
QY 3361 AGGAGAAATCTCTCTCTAGTGCCTGTGATGCCAAAGCCAGCCCTTCAAGAGTGTTC 3420  
DB 3361 AGGAGAAATCTCTCTCTAGTGCCTGTGATGCCAAAGCCAGCCCTTCAAGAGTGTTC 3420

QY 3421 CTTCTGTCTGCTTCCCTTTGTAATGTGATCCACTCGGGAAATGGTGTCTTACCCAGGA 3480  
DB 3421 CTTCTGTCTGCTTCCCTTTGTAATGTGATCCACTCGGGAAATGGTGTCTTACCCAGGA 3480  
QY 3481 AAAGTGCTTACCCAGTCAAGGGTCAACAGTGTGTGTGAGGATCAAAATGGCTCTGT 3540  
DB 3481 AAAGTGCTTACCCAGTCAAGGGTCAACAGTGTGTGTGAGGATCAAAATGGCTCTGT 3540  
QY 3541 GAAATACTGCAACCCATTTATTTGGTGTGATTTCTCAGGAGTCTGAATCTTCCCTCAGC 3600  
DB 3541 GAAATACTGCAACCCATTTATTTGGTGTGATTTCTCAGGAGTCTGAATCTTCCCTCAGC 3600  
QY 3601 AGTCTCTTCTCCAAACCCCTACCCAGAGCAACACTGGGATTTGAACCTTCTCTGACTC 3660  
DB 3601 AGTCTCTTCTCCAAACCCCTACCCAGAGCAACACTGGGATTTGAACCTTCTCTGACTC 3660  
QY 3661 TCTTCTTCCCTCAGTCTGACAACTAATGTCTCTCGGGACACCCAGCTAGGGCTTCTCC 3720  
DB 3661 TCTTCTTCCCTCAGTCTGACAACTAATGTCTCTCGGGACACCCAGCTAGGGCTTCTCC 3720  
QY 3721 CAACCTCCTTATCCAGCTGAATCTGGAATCTTCCAAACAGGGCTTGACCTGGGTCTGTT 3780  
DB 3721 CAACCTCCTTATCCAGCTGAATCTGGAATCTTCCAAACAGGGCTTGACCTGGGTCTGTT 3780  
QY 3781 GGTCCCACTGGCCAAACACACATCTTTGGCCAGATTTGGGATTTCTCAATAGATTTTATAGA 3840  
DB 3781 GGTCCCACTGGCCAAACACACATCTTTGGCCAGATTTGGGATTTCTCAATAGATTTTATAGA 3840  
QY 3841 CATTATTCTCCACAGACTTTTAAACATGGCTTGTCTTCCATACACATCCCGTCTAGA 3900  
DB 3841 CATTATTCTCCACAGACTTTTAAACATGGCTTGTCTTCCATACACATCCCGTCTAGA 3900  
QY 3901 TTTAAACATTTTATATACCAAGAAATTAACCAAGCAATAGAGTACTTTTCAGATATA 3960  
DB 3901 TTTAAACATTTTATATACCAAGAAATTAACCAAGCAATAGAGTACTTTTCAGATATA 3960  
QY 3961 AACTGTGTTTATATCTTATCTAGAGTGTCTGTATGTATAGGGGTATGACCTGGCTGA 4020  
DB 3961 AACTGTGTTTATATCTTATCTAGAGTGTCTGTATGTATAGGGGTATGACCTGGCTGA 4020  
QY 4021 AGTAATATTAACCATAGCTCTGGGAGGATTTACAGACCTTTTGCACTTTATGCTTTTTTG 4080  
DB 4021 AGTAATATTAACCATAGCTCTGGGAGGATTTACAGACCTTTTGCACTTTATGCTTTTTG 4080  
QY 4081 TGAACCTGATAACCATGGTCAATATTAAGCCAAATAACCTGGCAATTTCTGTGAATAAAC 4140  
DB 4081 TGAACCTGATAACCATGGTCAATATTAAGCCAAATAACCTGGCAATTTCTGTGAATAAAC 4140  
QY 4141 ATGCATATGTATCTTAAACCAAAACCAAAACCAAAAC 4175  
DB 4141 ATGCATATGTATCTTAAACCAAAACCAAAACCAAAAC 4175

## RESULT 2

US-10-188-832-21  
; Sequence 21, Application US/10188832  
; Publication No. US20040076955A1  
; GENERAL INFORMATION:  
; APPLICANT: Mack, David H.  
; APPLICANT: Aziz, Nataasha  
; APPLICANT: Eos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions  
; TITLE OF INVENTION: and Methods of Screening for Modulators of Bladder  
; TITLE OF INVENTION: Cancer  
; FILE REFERENCE: 018501-002330US  
; CURRENT APPLICATION NUMBER: US/10/188,832  
; CURRENT FILING DATE: 2002-11-22  
; PRIOR APPLICATION NUMBER: US 60/302,814  
; PRIOR FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: US 60/310,099  
; PRIOR FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: US 60/343,705  
; PRIOR FILING DATE: 2001-11-08

```
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2002-04-12
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 4202
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-188-832-21

Query Match      17.3%; Score 723.8; DB 16; Length 4202;
Best Local Similarity 84.9%; Pred. No. 6e-200;
Matches 823; Conservative 0; Mismatches 142; Indels 4; Gaps 1;

QY 10 GCTGCTCCGATGCTCCAGAGCGGCATGGGCGCCCGCCGACCTGGTGGGACCACTCGCGGGC 69
Db 52 GCTGCTCCGATGCTCCAGAGCGGCATGGGCGCCCGCCGACCTGGTGGGACCACTCGCGGGC 111
QY 70 TGGCAGTTCGAGGTGGATTGGTGGGAGGACCACTACACTATGCTGCTGCGCATTCGCGGA 129
Db 112 TGGTAGCTCGAGGTGGACTGGTGGGAGGACCACTACACTATGCTGCTGCTATCGCGGA 171
QY 130 GTTCTACACACGATCAGCAAGCTTTGTTTTCATTTTACCTCCCATCTGCATGCTT 189
Db 172 GTTCTACACACGATCAGCAAGCTTTTATTTTTCATTTTACCGCCCATCTGCATGCTT 231
QY 190 GTTCCGCGAGTACGCAAGCTTTTCAACAGCGGCATCTACTTATATGAGCGCTCTTAGT 249
Db 232 GTTTCGTCAGTATGCAACATGCTTCAACAGTGGCATCTACTTATCTGGACTCTTTTGGT 291
QY 250 TGTAGTGGGATGGATCTCTCTACTTCCATGCAAGCTGAGTTCTGCTGGTCAAGCT 309
Db 292 TGTAGTGGGATGGATCTCTCTACTTCCATGCAAGCTTGTCTTCTGGTCAAGCT 351
QY 310 TGATGAATCTGCCATCTGTGGGTTCTGATGTGTGCTTGGCCATGTGGTTTCCAGGAG 369
Db 352 TGATGAATCTGCCATCTGTGGGTTCTGATGTGTGCTTGGCCATGTGGTTTCCAGGAG 411
QY 370 GTATTTACCAAGATCTTTCCGAATGACAGGGCAGGTTCAAGGACGATGGTGTGTCT 429
Db 412 GTATTTACCAAGATCTTTCCGAATGACAGGGGAGTGTCAAGGAGTGGTGTGTCT 471
QY 430 GTCTGCAATTTACAGCTGCTTGGGTTTATCAAGCCCGCCATCAACAATTTTCCCTGAT 489
Db 472 GTCTGCGGTTTACAGCTGCTTGGGTTTATCAAGCCCGCCATCAACAATTTTCTCTGAT 531
QY 490 GATTCTGGGACTTCCATGCTGCTGCTTGTGTCAGAGCTGAAGAGGTGTGACAAATGT 549
Db 532 GACCTTGGGAGTTTCTTGGACTGCTGCTATCGCAGAGCTAAGAGGTGTGACAAAT 591
QY 550 GCGTGTGTTTAACTGGGCTCTTCTGCGCTCTGCTGAGCTCTGCTCTTCTTCTGCTG 609
Db 592 GCGTGTGTTTAACTGGGCTCTTCTGCGGCTCTGCTGAGCCCTGCGCCCTGCTTCTGCTG 651
QY 610 GATCAGCGACCAAGCTTCTGTCAGCTGCTCTCTCTTCACTTCCCTTACCTGCACTG 669
Db 652 GATCAGTACCGAGCTTCTGCGAGCTGCTGCTCTCTTCACTTCCCTTACCTGCACTG 711
QY 670 TGTGTGGCATTTCTCATCTGCTTCTGTCACCTGGGCTGTGTGCTTCCGCTTACTT 729
Db 712 CATGTGGCAGCATCTCATCTGCTTCTGCTACCTGGGCTGTGTGCTTTCGCTTACTT 771
QY 730 TGTATGCTGCTCAGAGTACTGAGCAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 789
Db 772 TGTATGCTGCTCAGAGTACTGAGCAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 831
QY 790 ATGGGCTTTTATTTGGTGTCTCTTATGTGCTCTTCTGTCGTCCTTCTGTCGTCCT 849
Db 832 ATGGGCTTTTATTTGGTGTCTCTTATGTGCTCTTCTGTCGTCCTTCTGTCGTCCT 891
QY 850 CAAGATCAGGTGATGGAAGGAGTGCAGGCTTCTTCTTACTTCTTCTTCTTCTTCTTCT 909
```

```
Db 892 CAAGATCAGGTGATGGCAAGATGGTGGCTTCTCTGCTTATCGCCCTC-----ATGC 947
QY 910 GCTGGGCTTCTGTTTGTAGCAAGATGCTGAGGGGTTGAGGAATTTGTTGTTGTTGGG 969
Db 948 AGTGGGCTTCTTGTGCTAGGAGACAGCCAGGAGTTCGAATAGTTGGGGTGTGGGCTA 1007
QY 970 TGTTTAAAA 978
Db 1008 TCITTTCAA 1016

RESULT 3
US-10-295-027-151
; Sequence 151, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 151
; LENGTH: 4212
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-151

Query Match      17.3%; Score 723.8; DB 15; Length 4212;
Best Local Similarity 84.9%; Pred. No. 6e-200;
Matches 823; Conservative 0; Mismatches 142; Indels 4; Gaps 1;

QY 10 GCTGCTCCGATGCTCCAGAGCGGCATGGGCGCCCGCCGACCTGGTGGGACCACTCGCGGGC 69
Db 62 GCTGCTCCGATGCTCCGAGTGGCCATGGGCGCCCGCCGACCTGGTGGGACCACTCGCGGGC 121
QY 70 TGGCAGTTCGAGGTGGATTGGTGGGAGGACCACTACACTATGCTGCTGCGCATTCGCGGA 129
Db 122 TGGTAGCTCGAGGTGGACTGGTGGGAGGACCACTACACTATGCTGCTGCTATCGCGGA 181
```

|    |  |      |   |      |
|----|--|------|---|------|
| Qy |  | 130  | GTTCTCAACACAGATCAGCAAGCTGTGTTTTTTCATTTTTACTGCCCATCTGCATGTGCTT     | 189  |
| Db |  | 182  | GTTCTCAACACAGATCAGCAATGTCTATTATTTTCATTTTACCGCCCATCTGCATGTGCTT     | 241  |
| Qy |  | 190  | GTTCCGGCAGTAGCAACAGTGCTTCAACAGCGGCATCTACTTTAATA TGGACGCTCCCTAGT   | 249  |
| Db |  | 242  | GTTTTCGT CAGTATGCAACATGCTTCAACAGTGGCATCTACTTAATCTGGACTCTTTTGGT    | 301  |
| Qy |  | 250  | TGTAGTGGGAATTGGATCTGTCTACTTTCCATPGCAACGCTGAGTTTCTGGGT CAGATGCT    | 309  |
| Db |  | 302  | TGTAGTGGGAATTGGATCCCGTCTACTTCCATPGCAACOCCTTAGTTTCTTGGGT CAGA TGCT | 361  |
| Qy |  | 310  | TGATGAACCTTGCCATTCTGTGGGTTCATGTGTGCTTTGGCCATGTGTTTCCCACGAG        | 369  |
| Db |  | 362  | TGATGAACCTTGCACTCCTTTGGGHTTCATGTGTGCTTTGGCCATGTGTTTCCCACAAG       | 421  |
| Qy |  | 370  | GTAFTTTACAAAGATCTTTCCGAAATGACAGGGCAGGTTCAAAGCAGTGTGTGTGTCCT       | 429  |
| Db |  | 422  | GTAFTTACAAAGATCTTTCCGAAATGACCGGGTAGGTTCAAAGTGTGGTCA GTGTCCT       | 481  |
| Qy |  | 430  | GTCCTGCAATTTACACGTGCTTGGCGTTTATCAAGCCGCCATCAACAATATTTCCCTGAT      | 489  |
| Db |  | 482  | GTCCTGGGTTACGACGTGCCCTGGCATTTGTCAAGCCTGCCATCAACAACATCTCTCTGAT     | 541  |
| Qy |  | 490  | GAFTCTGGGACTTCCATGCACTCGCGCTGCTTGTGTGCAGAGCTGAAGAGGTGTGCAATGT     | 549  |
| Db |  | 542  | GACCTTGGGAGTTCCTTGCACCTGCACCTCATCGCAGAGCTAAGAGGTGTGCAACAT         | 601  |
| Qy |  | 550  | CGGTGTGTTAAGCTTGGGCTCTTCTCTGGGCTCTGTGTGAGACTGTGGCTCTCTTCTGCTG     | 609  |
| Db |  | 602  | CGGTGTGTTAAGCTTGGGCTCTTCTCGGGCTCTGTGTGAGACCCTGGCCCTGTGTTCTGCTG    | 661  |
| Qy |  | 610  | GATCAGCGACCAAGCCTTCTGTGAGCTGCTCTCTCTTTCACATTCCCTACCTGCACTG        | 669  |
| Db |  | 662  | GATCAGTGACCGAGCTTCTGCGAGCTGTGTGTCAITTCCTTCCCTTACCTTGCACTG         | 721  |
| Qy |  | 670  | TGTTGGCATATTCTCATCTGCTTGTCTGTACTCTGGCTGTGTGTCTTCGCTACTT           | 729  |
| Db |  | 722  | CATGTGCCACATCCTCATCTGCTTGTGCTTACCTGGGCTGTGTATGCTTTTGCTACTT        | 781  |
| Qy |  | 730  | TGATGCTGCTCAGAGATACCTGAGCAAGGTTCAGTCA TCAGATTCTGGCCCCACGGAAA      | 789  |
| Db |  | 782  | TGATGCTGCTCAGAGATTCCTGAGCAAGGCCCTGT CATCAAGTTCTTGGCCCAATGAAA      | 841  |
| Qy |  | 790  | ATGGGCTTTTATTGGGTGCCCTTATGTGTCCCTTCTGTGTGCCACAAGAGTCCCA GT        | 849  |
| Db |  | 842  | ATGGGCTTTCATTGGTGTCCCTATGTCTCTCTGTGTGCCACAAGAAATCATCAGT           | 901  |
| Qy |  | 850  | CAAGATCACGTGATGGCAGCGATGACCGACTCTCTATCTTATTTTCGATGGCGGC           | 909  |
| Db |  | 902  | CAAGATCACGTGATGGCAAGATGGTGGCTGCTTCTCTGCTTATTCGCCCTC --- ATGC      | 957  |
| Qy |  | 910  | GCTGGGCTTCGTTTGCTAGCAAGATGGCTCAGGGGGTTGAGGAATTTGGTGTGTGTGGG       | 969  |
| Db |  | 958  | AGTGGGCTTCCTTTTGCTAGGAAGACAGCCCAAGGGAGTTCGAATAGTTGGGTGTGGGCTA     | 1017 |
| Qy |  | 970  | TGTTTAAAA   | 978  |
| Db |  | 1018 | TCCTTTCAA   | 1026 |

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 828
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(825)
US-10-017-410-3

```

| Query Match           | 15.9%; | Score 564.8;  | DB 13;          | Length 828;       |
|-----------------------|--------|---|-----------------|-------------------|
| Best Local Similarity | 87.7%; | Pred. No. 3.8e-183;   |                 |                   |
| Matches               | 726;   | Conservative 0;   | Mismatches 102; | Indels 0; Gaps 0; |
| Qy                    | 35     | ATGGGCCCCGGCACTGGTGGGACCACTGCGGGCTGGCAGTTTCGGAGGTGGATTTGGTGC  | 94              |                   |
| Db                    | 1      | ATGGGCCCCGGCACTGGTGGGACCACTGCGGGCTGGCAGTTTCGGAGGTGGATTTGGTGC  | 60              |                   |
| Qy                    | 95     | GAGGACAACTACACTATCTGTCGCTCGCAATTCGCCGAGTTCACAAACGATCAGCAAGTGC | 154             |                   |
| Db                    | 61     | GAGGACAACTACACATCTGTCGCTCGTATCGCCGAGTTCACAAACGATCAGCAAGTGC    | 120             |                   |
| Qy                    | 155    | TTGTTTTTCATTTTACCTCCCATCTGCATGTGCTTTTCGCCAGTAGCAAACTGCTTC     | 214             |                   |
| Db                    | 121    | TTATTTTTTCATTTTACGCCCACTGCAATGTGCTTTGTTGATGAGTATGCAATGCTTG    | 180             |                   |
| Qy                    | 215    | AACAGCGGCATCTACTTAAATATAGACGCTCCTAGTTGATGAGGGAATGGAATCTGTCTAC | 274             |                   |
| Db                    | 181    | AACAGTGACATCTACTTAAATCTGGACTCTTTTGGTTGTAGTGGAAATGGATCCGCTAC   | 240             |                   |
| Qy                    | 275    | TTCCATGCACAGCTGAGTTTCTGGGTGAGATGCTTGATGAATCTGCATCTCTGCGGTT    | 334             |                   |
| Db                    | 241    | TTCCATTTTACCCCTAGTTTCTGGGTGAGATGCTTGATGAATCTGAGTCTCTTTGGGTT   | 300             |                   |
| Qy                    | 335    | CTGATGTGTGTTTGGCCATGTGTTTCCAGGAGGTATTTACCAAGATCTTTTCGGAAT     | 394             |                   |
| Db                    | 301    | CTGATGTGTGTTTGGCCATGTGTTTCCAGGAGGTATCTACCAAGATCTTTTCGGAAT     | 360             |                   |
| Qy                    | 395    | GACAGGGGAGGTTCAAGGCAAGTGTGTGTCTGTCTGCAATTAACAAGTGTGTCGCG      | 454             |                   |
| Db                    | 361    | GACAGGGGTAGGTTCAAGGTGGTGGTCAGTGTCTGTCTGCGGTTTACGACGTGCTGSCA   | 420             |                   |
| Qy                    | 455    | TTTATCAAGCCCGCCATCAACAATTTTCCCTGATGATTTCTGGGACATTCACATGCAC    | 514             |                   |
| Db                    | 421    | TTTGTCAAGCCTGCGCATCAACAATCTCTCTGATGACCCCTGGGAGTTCCTTTGACTGCA  | 480             |                   |
| Qy                    | 515    | CTGCTTGTTCAGAGCTGAAGAGGTGACAAATGTGTGTGTTTAAAGCTGGGCTCTTTC     | 574             |                   |
| Db                    | 481    | CTGCTCATCGCAGAGCTAAAGAGGTGACAAACATGCGTGTGTTTAAAGCTGGGCTCTTTC  | 540             |                   |
| Qy                    | 575    | TCTGGCTCTGTGTGACTCTGGCTCTCTCTGCTGGATCAGCAGCAAGCCCTCTGTGAG     | 634             |                   |
| Db                    | 541    | TCGGGCTCTGTGTGGACCTGGCCCTGTTCTGTCTGGATCAGTACCGAGCTTCTGTGGAG   | 600             |                   |
| Qy                    | 635    | CTGCTCTCTCTTTTCACTTCCCTACCTGCACGTGTGTGTGGCAATATTCATCTGCCTT    | 694             |                   |
| Db                    | 601    | CTGCTGTATCTTCAACTTCCCTACCTGCACGTGTGTGGCAATATTCATCTGCCTT       | 660             |                   |
| Qy                    | 695    | GCTTGTACCTGGGCTGTGTGTGCTTTCGCTTACTTTGATGCTGCTCAGAGATACCTGAG   | 754             |                   |
| Db                    | 661    | GCTGCCCTACCTGGGCTGTGTATGCTTTTGGCTACTTTTGTGCTCTCAGAGATTCCTGAG  | 720             |                   |
| Qy                    | 755    | CAAGGTCAGTTCATCAGATTCGCGCCAGGAGAAATGGGCTTTTATTGGTGTCCCTTAT    | 814             |                   |
| Db                    | 721    | CAAGGCCCTGTTCATCAAGTTCTGCCCCAATGAGAAATGGGCTTCTATTGGTGTCCCTAT  | 780             |                   |
| Qy                    | 815    | GTGTCCCTTCTGTGTGCCCAAGAGATGTGCCAGTCAAGATCAAGTGA               | 862             |                   |
| Db                    | 781    | GTGTCCCTCTGTGTGCCCAAGAGAAATTCATCAGTCAAGATCAAGTGA              | 828             |                   |

RESULT 5  
US-10-182-447-6

```

RESULT 4
US-10-017-410-3
; Sequence 3, Application US/10017410
; Publication No. US20020115094A1
; GENERAL INFORMATION:
; APPLICANT: Farnham, Peggy J
; APPLICANT: Graveel, Carrie R
; TITLE OF INVENTION: Polynucleotide Differentially Expressed in Liver Cancer
; FILE REFERENCE: 960296..97401
; CURRENT APPLICATION NUMBER: US/10/017,410
; CURRENT FILING DATE: 2001-12-14
; NUMBER OF SEQ ID NOS: 4

```







|      |    |  |      |
|------|----|--|------|
| 3701 | QY | -----CACCCAGCTAGG-- -CCTTCCCAACTCCTTATPC-----AGCTGA            | 3739 |
|      |    |  |      |
| 652  | Db | TGGATGCTGAAGCTCACCCAGTCAGGCGCCCTCTCCTAGCTCCTTTTACACTGAAATTAA   | 593  |
|      |    |  |      |
| 3740 | QY | ACTTGGATTTCTTCCCAACGAGGGTTGACCTGGGTGCTGTTGGTCCACCTGGCCACAAC    | 3799 |
|      |    |  |      |
| 592  | Db | TCTGAAGCTTTTCATAGCCAAGGCTTTGCCCTAGGTGCTATTATTCCAGCTGGCCAAAGAG  | 533  |
|      |    |  |      |
| 3800 | QY | ACATCTTTGGCCAGATTTGGGATTTCTCAATAGATTTTATAGACATATTCTCCACACAGCT  | 3859 |
|      |    |  |      |
| 532  | Db | AAGTCTTTGGCCAGATTTGGGATTTCTCAATGGATTTTATAGACATAATTCCTCTGCAAACT | 473  |
|      |    |  |      |
| 3860 | QY | TTAAAAACA-----TGGCTTGTGCTCTTCCATAC                             | 3887 |
|      |    |  |      |
| 472  | Db | TAAAAAATAAATAAACCCCTACTTTATAGACATAATTGTTTGAATTGTAICTTCTCTGT    | 413  |
|      |    |  |      |
| 3888 | QY | ACATCCGGTCAGATTTTAAAACTATTTTATAACCAACAGAAATTAAACCAAGCAATAGAGT  | 3947 |
|      |    |  |      |
| 412  | Db | ATGTTAAACACGATTTTAAAACTATTTTATAACCACAATATGTAATCAGACA-ATATAGT   | 354  |
|      |    |  |      |
| 3948 | QY | ACTTTCAGATATAAACHTGTGTTTCATACTTTATGTAGAGTGTGCTATGTTATAGCCGTAT  | 4007 |
|      |    |  |      |
| 353  | Db | GTTTTTCAGATATAACCTTGTTTATACCTTATGTAG-GTGTCCCATATAAGGGTGGCAT    | 295  |
|      |    |  |      |
| 4008 | QY | GTACCCCTGGCTGAAGTAATAATTAAC---CATAGCTCTGGGAGATTTACAGAC-----    | 4057 |
|      |    |  |      |
| 294  | Db | GCCCACTGGCTGTGGTAAATTTTAAATCCTCATCTGTTGGAGTGACCTTAAGGCCTTTTG   | 235  |
|      |    |  |      |
| 4058 | QY | -----CTTTTGCATTTTANGCTTTTT--TGTGAACTCTGATAACCACTGCTCAATATTA    | 4108 |
|      |    |  |      |
| 234  | Db | AAGTGGAGCTTTGCACTTTTATACCTTTTCTGTGAACTAATGATPAACTATATTTCATATTA | 175  |
|      |    |  |      |
| 4109 | QY | AAGCCAATAACGTGCAATTTTCTGTGAATAAACAATGCATATGATCT                | 4154 |
|      |    |  |      |
| 174  | Db | AAGCTG-TAAGTGGCATTTTCAGCAATGAATATGTACATCTTTGT                  | 130  |

```

RESULT 11
US-10-106-698-611
; Sequence 611, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 611
; LENGTH: 1473
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-106-698-611

Query Match          3.0%; Score 125.8; DB 15; Length 1473;
Best Local Similarity 62.4%; Pred. No. 4e-25;
Matches 440; Conservative 0; Mismatches 177; Indels 88; Gaps 12;

QY      3527 AAACATGGCTCTGTGAAATATCTGCCACCCCATTTCTATTGGTTGGATTTCTCAGGAGCTCG 3586
Db      637 ACACAAGGCTTTGTAAAAATACGACCACCTATTCCACTTACTGGATCTGTCAGGTGTGTA 696

QY      3587 AATCTTCCCTACAGAGTCCTCTTCTCCCAACCCCTACCCAGAGCCACACTGGGATTTGA 3646
Db      697 AAACCTTCTCTGCCAG-----TTATCATGCTGTTCCATGAGCCCTCAGGACTCGGATTTGA 751

```

|      |    |            |                                   |                                    |            |      |
|------|----|------------|-----------------------------------|------------------------------------|------------|------|
| 3647 | Qy | ACCTTCTGAC | CTCTCTTCTTCCGCTCAGTCTGAC          | CAACTAAATGGTCTCTGGGA               | -----      | 3700 |
| 752  | Db | GCCTTCCTGG | ---CTCTTTATCCCTTGGGGCAGACATGGA    | ACCATCTCTGAGGGACACAGG              | 808        |      |
| 3701 | Qy | -----      | CACCCAGCTAGGG                     | ---CCTTCCCAACTCCTTATCC             | ---AGCTGAA | 3740 |
| 809  | Db | TGGATGCTGA | AGCTCACCAGCTCAGGGCCCTCTCCTAGCT    | CCTTTTACACTGCAATTTAA               | 868        |      |
| 3741 | Qy | CTTGGATTCT | CCCAACCAAGGGCTTCACTGGGTGCTGTGGT   | CCCACTGGGCCAACAA                   | 3800       |      |
| 869  | Db | TCTGAAGCTT | CATAGCCAAGGCTTGTAGGTCTATTATTTCC   | AGCTGGCCAAAGAGA                    | 928        |      |
| 3801 | Qy | CATCTTTGGC | AGATTGGGATTCCTCAATAGATTTTATAGAC   | ATTATCTCCACACACTT                  | 3860       |      |
| 929  | Db | AGTCTTTGGG | CCAGATTGGGATTTCTCAATGGATTTTATAGAC | ATAATTTCCCTCGCAAACTT               | 998        |      |
| 3861 | Qy | TAAAACA    | -----                             | -----TGGCTGTGTCTTTCCCATACA         | 3888       |      |
| 989  | Db | AAAAAATAA  | ATAACCCCTACTTTATAGGACTTAAT        | TGTTTGAATGTATCTTCTCTGTA            | 1048       |      |
| 3889 | Qy | CATCCGGTCA | GATTTTAAAACTATTTTATAACCA          | CAGGAATTTAAACCAAGCAAAATAGAGTA      | 3948       |      |
| 1049 | Db | TGTTAAACCA | GATTTTAAAACTATTTTATAACCA          | CAATATGTAATCAGAGCA-ATATAGTG        | 1107       |      |
| 3949 | Qy | CTTTCAGATA | ATAAAGTGTGTTTCATCTTATGTAGAGTG     | TGCTATGTATAGCGGATG                 | 4008       |      |
| 1108 | Db | TTTTTCAGAT | ATATACCTTGTTTTATACCTTATGTAG       | -GTGTCTACATAAGGGTGGCATG            | 1166       |      |
| 4009 | Qy | TACCTGGCTG | AGTAATAATTAC                      | ---CATAGCTCTGGGAGGATTTACAGAC       | -----      | 4057 |
| 1167 | Db | CCCACCTGG  | CTGTGTAAAAATTTAATCCTCAAT          | TGCTTTGGGAGTGACTTAAGGCGCTTTTGA     | 1226       |      |
| 4058 | Qy | -----      | CTTTTGGCACTTTATGCTTTTTT           | -TGTGAACCTCTGATAACCAATGGTCAATATTAA | 4109       |      |
| 1227 | Db | AGTGGAGCT  | TTTGCACTTTATATCTTTTCTGTGA         | CACTATGATAAATTTTGATATTAA           | 1286       |      |
| 4110 | Qy | AGCCAATA   | CTGGCATTTCTGTGGAATAAACATGCATATG   | TATCT 4154                         |            |      |
| 1287 | Db | AGCTG-TAA  | GTCATTTTTCAGCAAAATGAATATG         | TATCATCTTTGT 1330                  |            |      |

```

RESULT 12
US-10-242-535A-50197
; Sequence 535A-50197, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 50197
; LENGTH: 418
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-50197

Query Match      1.6%; Score 65; DB 16; Length 418;
Best Local Similarity 55.4%; Pred. No. 1,1e-07;
Matches 201; Conservative 0; Mismatches 140; Indels 22; Gaps

QV      3527 AAACATGGCTCTGTGAAATACTGCCACCCTATTCTATTGGTTGGATTCTTCAGAGTCTC

```

Db 27 ACACAGGCTTTGTAATAATACGACCACTATTTCACCTTACTGGATCTGTGAGGTGTA 86  
QY 3587 AATCTTCCCTCAGGAGCTCTTCTCCCAACCCCTACCCAGAGCCAACTGGGATTTGA 3646  
Db 87 AAACCTTCTCGCCAGTTTCATCATGCTTCCATGAGCCCTCAGGACTGGGATTTGAGCCTT 146  
QY 3647 ACCTTCTCTGACTCTCTTCTTCCCTCAGGCTGTGACAACTAAT-----GGTCTCTGG 3697  
Db 147 CTGGCTCTTTATCCCTTGGGGCAGGACATGGAACCATCTCTGAGGACCCAGGTGGATGC 206  
QY 3698 GGACACCCAGCTAGGCGCTTCCCAACTCCTTATCCAGCTGAACCTTGGATCTTCCCAA- 3756  
Db 207 TGAAGCTCACCCAGTTCAGGGGCCCTCTCTAGCTCTCTTTTACACTGAAATTAATCTGAA 266  
QY 3757 -----CCAGGGCTTCCACTGGGTGCTGTGGTCCCACTGGCCCAACACACATCT 3805  
Db 267 GCTTTCATAGCAGGCTTCCCTAGGCTGTCTATTTCAGCTGGCCAAAGAGA-AGCT 325  
QY 3806 TTGGCCAGATTGGGATTTCTAATAGATTTTATAGACATTTATCTCCACAGACTTTAAAA 3865  
Db 326 TGGGCCAGATTGGGATTTCTAATGATTTTATAGACATTTATCTCCCTGCAAACTTTAAAA 385  
QY 3866 CAT 3868  
Db 386 AAT 388

## RESULT 13

US-10-085-783A-50197  
; Sequence 50197, Application US/10085783A  
; Publication No. US20040037841A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2002  
; CURRENT APPLICATION NUMBER: US/10/085,783A  
; CURRENT FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 50197  
; LENGTH: 418  
; TYPE: DNA  
; ORGANISM: Human  
US-10-085-783A-50197

Query Match 1.6%; Score 65; DB 16; Length 418;  
Best Local Similarity 55.4%; Pred. No. 1.1e-07;  
Matches 201; Conservative 0; Mismatches 140; Indels 22; Gaps 3;

QY 3527 AACATGGCTCTGTGAAATACCTGCCACCCATCTATTGGTTGGATTTCTCAGGATCTG 3586  
Db 27 ACACAGGCTTTGTAATAATACGACCACTATTTCACCTTACTGGATCTGTGAGGTGTA 86  
QY 3587 AATCTTCCCTCAGGAGCTCTTCTCCCAACCCCTACCCAGAGCCAACTGGGATTTGA 3646  
Db 87 AAACCTTCTCTCGCCAGTTTCATCATGCTTCCATGAGCCCTCAGGACTGGGATTTGAGCCTT 146  
QY 3647 ACCTTCTCTGACTCTCTTCTTCCCTCAGGCTGTGACAACTAAT-----GGTCTCTGG 3697  
Db 147 CTGGCTCTTTATCCCTTGGGGCAGGACATGGAACCATCTCTGAGGACCCAGGTGGATGC 206  
QY 3698 GGACACCCAGCTAGGCGCTTCCCAACTCCTTATCCAGCTGAACCTTGGATTTCTCCCAA- 3756  
Db 207 TGAAGCTCACCCAGTTCAGGGGCCCTCTCTAGCTCTCTTTTACACTGAAATTAATCTGAA 266

QY 3757 -----CCAGGGCTTGCCTGGGTGCTGTGGTCCCACTGGCCAAACACATCT 3805  
Db 267 GCTTTCATAGCAGGCTTTCCTCCTAGGCTGTCTATTTCAGCTGGCCAAAGAGA-AGCT 325  
QY 3806 TTGGCCAGATTGGGATTTCTAATAGATTTTATAGACATTTATCTCCACAGACTTTAAAA 3865  
Db 326 TGGGCCAGATTGGGATTTCTAATGATTTTATAGACATTTATCTCCCTGCAAACTTTAAAA 385  
QY 3866 CAT 3868  
Db 386 AAT 388

## RESULT 14

US-10-322-281-265/c  
; Sequence 265, Application US/10322281  
; Publication No. US20040126762A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc S. Malandro  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: 529452001000  
; CURRENT APPLICATION NUMBER: US/10/322,281  
; CURRENT FILING DATE: 2002-12-17  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 265  
; LENGTH: 77530  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(77530)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-322-281-265

Query Match 1.5%; Score 63.6; DB 17; Length 77530;  
Best Local Similarity 63.2%; Pred. No. 9.2e-06;  
Matches 115; Conservative 0; Mismatches 64; Indels 3; Gaps 1;

QY 2701 GGGGCTGGAGATTGGTTCAACAGTTAAGAACCCGTTGCTCTCTCCACAGACGACCCAG 2760  
Db 30410 GAGGCTGGAGAGAGGCTCATCAGTTAGGAGCACTGGCTCTCTCTGAGGGCCCAACA 30351  
QY 2761 CGACTTGTGAGGCTCTATAACCA---GCAACTCAGCCCTGGGCACTGTGAGCCCTCTTC 2817  
Db 30350 CTCACATGGTGGCTCACACCACTCCCACTCCCAAGAGATATGATGCCCTCTTC 30291  
QY 2818 TGACTTCTAAGGGCAATTGGTGACACGTCATACACGAGGCAAGCAAAACATTTATACATGT 2877  
Db 30290 TAGCCTCCGACACACACAGGGGTACATACATACAGCAAGCAGAACATTCATACAT 30231  
QY 2878 AA 2879  
Db 30230 AA 30229

## RESULT 15

US-09-771-208-20/c  
; Sequence 20, Application US/09771208  
; Patent No. US20020155564A1  
; GENERAL INFORMATION:  
; APPLICANT: MEDRANO, JUAN  
; APPLICANT: BRADFORD, ERIC  
; APPLICANT: HORVAT, SIMON  
; TITLE OF INVENTION: CLONING OF A HIGH-GROWTH GENE  
; FILE REFERENCE: 407T-923710US  
; CURRENT APPLICATION NUMBER: US/09/771,208  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: US 08/999,477  
; PRIOR FILING DATE: 1997-12-29  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.0



```
; SEQ ID NO 20
; LENGTH: 659158
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (123459)..(123478)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (602466)..(602485)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (546998)..(547017)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (494715)..(494814)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (390986)..(391005)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (346860)..(346883)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (317174)..(317193)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (280353)..(280373)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (271829)..(271848)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (183872)..(183891)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (170625)..(170645)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; LOCATION: (132680)..(132700)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc feature
; OTHER INFORMATION: n is a, c, g, or t
;
Query Match      1.5%; Score 61.4; DB 9; Length 659158;
Best Local Similarity 60.9%; Pred. No. 0.00017;
Matches 137; Conservative 0; Mismatches 81; Indels 7; Gaps 2;

Qy 2699 CTGGGCTGGAGAGTTGTTCAACAGTTAAGAACCCGTGTGCTCTCTCCAGACGACCC 2758
Db 318008 CTAGGGCTGGTGAGATGGCTCAGCAGTTAAGAGCACTTGCTGCTTCTCTAAGAACCCAC 317949

Qy 2759 AGCGACTTGTGAGGCTCATACACGCA---ACTCCAGCCCTGGGGCATCTGACGCCCTC 2814
Db 317948 ACCTACCTATGTGGTGGCTAATACAGACACTCCGGTCCAGGGGATGTGATGCCCTC 317889

Qy 2815 TTCTGACTTCTAAGGCCATTGGTGCACA---CGTACATACAGGAGGCAAAACATTAT 2871
Db 317888 TTCTGGCTCTCAGGGGCACTGCATACACATGTGTGCACAGGTATGCAGAGGAATATCCAC 317829

Qy 2872 ACATCTAAGTAATAAATGCAATAGTTAGTGACGGCTGAAGGA 2916
Db 317828 ACACATAAAATTTAAATCAAAATATTTCTTAGGAGAGGGAGAGA 317784
```

Search completed: November 21, 2004, 07:43:59  
Job time : 1957.22 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 20:50:47 ; Search time 11519.4 Seconds  
(without alignments)  
13206.895 Million cell updates/sec

Title: US-10-017-410-1

Perfect score: 4175

Sequence: 1 ggcacagggtctcgat.....aaaaaaaaaaaaaaaaaac 4175

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_hic.\*

4: gb\_est3.\*

5: gb\_est4.\*

6: gb\_est5.\*

7: gb\_est6.\*

8: gb\_gssi.\*

9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID      | Description        |
|------------|--------|-------------|--------|------------|--------------------|
| 1          | 3028.8 | 72.5        | 3071   | 3 AK035314 | AK035314 Mus muscu |
| 2          | 1341   | 32.1        | 1341   | 3 BC023423 | BC023423 Mus muscu |
| 3          | 780    | 18.7        | 1022   | 5 BU511164 | BU511164 AGENCOURT |
| 4          | 714.4  | 17.1        | 717    | 7 CN537424 | CN537424 UI-M-HS0- |
| 5          | 681.2  | 16.3        | 688    | 2 BB108761 | BB108761 BB108761  |
| 6          | 681    | 16.3        | 681    | 9 AY401891 | AY401891 Mus muscu |
| 7          | 670.6  | 16.1        | 1388   | 6 CF110927 | CF110927 Shultzomi |
| 8          | 668.4  | 16.0        | 670    | 7 CN721804 | CN721804 E0830A03- |
| 9          | 667.6  | 16.0        | 1173   | 3 AK085306 | AK085306 Mus muscu |
| 10         | 662.6  | 15.9        | 955    | 1 AV227941 | AV227941 AV227941  |
| 11         | 653.6  | 15.7        | 868    | 6 CA976684 | CA976684 AGENCOURT |
| 12         | 631.4  | 15.1        | 797    | 6 CA463294 | CA463294 AGENCOURT |
| 13         | 616.2  | 14.8        | 651    | 2 BB627599 | BB627599 BB627599  |
| 14         | 594    | 14.2        | 617    | 5 BW942111 | BW942111 UI-M-CG0p |
| 15         | 576.8  | 13.8        | 606    | 8 AZ411514 | AZ411514 IM018407  |
| 16         | 575.2  | 13.8        | 1124   | 4 BI080821 | BI080821 602878593 |
| 17         | 568.2  | 13.6        | 916    | 2 BF144307 | BF144307 601787125 |
| 18         | 563.6  | 13.5        | 633    | 2 BB660847 | BB660847 BB660847  |
| 19         | 559.4  | 13.4        | 562    | 2 AW701866 | AW701866 uq95909.x |
| 20         | 552    | 13.2        | 552    | 2 AW490555 | AW490555 UI-M-BH3- |
| 21         | 548    | 13.1        | 548    | 5 BX523075 | BX523075 BX523075  |
| 22         | 543.4  | 13.0        | 681    | 9 AY401890 | AY401890 Pan trogl |
| 23         | 541.8  | 13.0        | 681    | 9 AY401889 | AY401889 Homo sapi |
| 24         | 497.2  | 11.9        | 731    | 6 CA503516 | CA503516 UI-R-F00- |

|    |       |      |      |   |          |           |
|----|-------|------|------|---|----------|-----------|
| 25 | 497   | 11.9 | 565  | 6 | CF115220 | Shultzomi |
| 26 | 480   | 11.5 | 480  | 2 | BF662488 | maab6c03. |
| 27 | 479.6 | 11.5 | 1072 | 5 | BQ715126 | AGENCOURT |
| 28 | 476.6 | 11.4 | 548  | 6 | CF169369 | B0812G07- |
| 29 | 475   | 11.4 | 492  | 1 | AA212916 | mw86f03.r |
| 30 | 474   | 11.4 | 490  | 1 | AA212925 | mw86h04.r |
| 31 | 471.2 | 11.3 | 528  | 2 | BB764116 | BB764116  |
| 32 | 470.8 | 11.3 | 485  | 1 | AA475320 | vh14a03.r |
| 33 | 470   | 11.3 | 505  | 2 | BB752973 | BB752973  |
| 34 | 465   | 11.1 | 690  | 4 | BM387041 | UI-R-CN1- |
| 35 | 440   | 10.5 | 440  | 2 | BF661433 | maab6c03. |
| 36 | 437.8 | 10.5 | 442  | 2 | BF468547 | UI-M-BH3- |
| 37 | 431.8 | 10.3 | 448  | 1 | AA462019 | VG5d02.r  |
| 38 | 425.6 | 10.2 | 432  | 6 | CF169808 | B0818D08- |
| 39 | 419.2 | 10.0 | 622  | 6 | CB723138 | UI-M-GH0- |
| 40 | 417.2 | 10.0 | 422  | 4 | BG230006 | mac25g04. |
| 41 | 412.2 | 9.9  | 769  | 6 | CA510569 | UI-R-FJ0- |
| 42 | 410   | 9.8  | 427  | 2 | AW702104 | uq95g09.y |
| 43 | 408   | 9.8  | 408  | 1 | AA462654 | vg58e01.r |
| 44 | 405   | 9.7  | 498  | 4 | BI848265 | 470659 MA |
| 45 | 400.4 | 9.6  | 595  | 1 | AA848826 | EST191587 |

## ALIGNMENTS

|            |  |             |      |        |                 |
|------------|--|-------------|------|--------|-----------------|
| RESULT 1   | AK035314   | 3071 bp     | mrna | linear | HTC 03-APR-2004 |
| LOCUS      | Mus musculus adult male urinary bladder cDNA, RIKEN full-length enriched library, clone:9530014821 product:similar to CANCER RELATED GENE-LIVER 1 [Mus musculus], full insert sequence.  |             |      |        |                 |
| DEFINITION | AK035314   |             |      |        |                 |
| ACCESSION  | AK035314   |             |      |        |                 |
| VERSION    | AK035314.1   | GI:26084521 |      |        |                 |
| KEYWORDS   | HTC; CAP trapper.  |             |      |        |                 |
| SOURCE     | Mus musculus (house mouse)   |             |      |        |                 |
| ORGANISM   | Mus musculus   |             |      |        |                 |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  |             |      |        |                 |
| AUTHORS    | Carninci, P. and Hayashizaki, Y.   |             |      |        |                 |
| TITLE      | High-efficiency full-length cDNA cloning   |             |      |        |                 |
| JOURNAL    | Mech. Enzymol. 303, 19-44 (1999)   |             |      |        |                 |
| MEDLINE    | 99279253   |             |      |        |                 |
| PUBMED     | 10349636   |             |      |        |                 |
| REFERENCE  | 2  |             |      |        |                 |
| AUTHORS    | Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.   |             |      |        |                 |
| TITLE      | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes   |             |      |        |                 |
| JOURNAL    | Genome Res. 10 (10), 1617-1630 (2000)  |             |      |        |                 |
| MEDLINE    | 20499374   |             |      |        |                 |
| PUBMED     | 11042159   |             |      |        |                 |
| REFERENCE  | 3  |             |      |        |                 |
| AUTHORS    | Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Ishiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. |             |      |        |                 |
| TITLE      | RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer   |             |      |        |                 |
| JOURNAL    | Genome Res. 10 (11), 1757-1771 (2000)  |             |      |        |                 |
| MEDLINE    | 20530913   |             |      |        |                 |
| PUBMED     | 11076861   |             |      |        |                 |
| REFERENCE  | 4  |             |      |        |                 |
| AUTHORS    | The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.   |             |      |        |                 |
| TITLE      | Functional annotation of a full-length mouse cDNA collection   |             |      |        |                 |
| JOURNAL    | Nature 409, 685-690 (2001)   |             |      |        |                 |
| MEDLINE    |  |             |      |        |                 |
| PUBMED     |  |             |      |        |                 |
| REFERENCE  | 5  |             |      |        |                 |
| AUTHORS    | The FANTOM Consortium and the RIKEN Genome Exploration Research  |             |      |        |                 |

## Group Phase I &amp; II Team.

Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs

JOURNAL  
REFERENCE  
AUTHORS

Nature 420, 563-573 (2002)  
6 (bases 1 to 3071)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayashida, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komori, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.

## Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,  
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/  
URL: http://phantom.gsc.riken.jp/.

## FEATURES

## source

1. 3071

Location/Qualifiers

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM DB:9530014B21"

/db\_xref="taxon:10090"

/clone="9530014B21"

/sex="male"

/tissue\_type="urinary bladder"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="adult"

## misc\_feature

1. 3071

/notes="similar to CANCER RELATED GENE-LIVER 1 [Mus

musculus] (SPTRIAL40408, evidence: FASTY, 100%ID,

77.8%length, match=642)"

## ORIGIN

Query Match 72.5%; Score 3028.8; DB 3; Length 3071;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 3065; Conservative 0; Mismatches 2; Indels 5; Gaps 3;  
QY 1065 AGTATCCTTTCCAGGCATAGCGGCTGCTGGGTTATGTCATATGACATCAACAGTA 1124  
DB 2 AGTATCCTTTCCAGGCATAGCGGCTGCTGGGTTATGTCATATGACATCAACAGTA 61  
QY 1125 GTTCTTTGGAGACCTAGGCAACCCAAAGTTTCTTCTGGAGAGGCTAGCTTGTCTT 1184  
DB 62 GTTCTTTGGAGACCTAGGCAACCCAAAGTTTCTTCTGGAGAGGCTAGCTTGTCTT 121  
QY 1185 CCACCATTTCCAGACTCTCAACCCCTGATAACAGACCTTCTGATTTGGTGATGAAG 1244  
DB 122 CCACCATTTCCAGACTCTCAACCCCTGATAACAGACCTTCTGATTTGGTGATGAAG 181  
QY 1245 GTTCCAGAACTTTTCATTTTCCGGGAGAACTGTCTTCAACAAACCAAGTGGGCAAA 1304  
DB 182 GTTCCAGAACTTTTCATTTTCCGGGAGAACTGTCTTCAACAAACCAAGTGGGCAAA 241  
QY 1305 ACACGTGTGGGGGTGTGACTAAGACGGGTGGTGTCTACCTCGCACTCTTACCTTCCC 1364

|    |      |  |      |
|----|------|--|------|
| QY | 2442 | GCTTAAATATGTTTAAATGACCTTGGAGTGAATGTTTGGCAACCTTAGGGGTTTATG  | 2501 |
| Db | 1380 | GCTTAAATATGTTTAAATGACCTTGGAGTGAATGTTTGGCAACCTTAGGGGTTTATG  | 1439 |
| QY | 2502 | GATCAGAAATGCGGCTGAGTCCCTTGGTCTTGGAGTCTAGAGTTTTCAGAGGCAAAAT | 2561 |
| Db | 1440 | GATCAGAAATGCGGCTGAGTCCCTTGGTCTTGGAGTCTAGAGTTTTCAGAGGCAAAAT | 1499 |
| QY | 2562 | CAAACTAGCCCTTGTCTGCTTCACTTACAGGCTCGGAGCTGCCCTTATATATCTA    | 2621 |
| Db | 1500 | CAAACTAGCCCTTGTCTGCTTCACTTACAGGCTCGGAGCTGCCCTTATATATCTA    | 1559 |
| QY | 2622 | CATAAGGACTTATATACATAAGTCTGTATAAATGCTCTGAAGATGACCTAGTGCCT   | 2681 |
| Db | 1560 | CATAAGGACTTATATACATAAGTCTGTATAAATGCTCTGAAGATGACCTAGTGCCT   | 1619 |
| QY | 2682 | TCATCTGGAAGGTCGCTCGGGCTGAGAGTGTGTTCAACAGTTAAGAACCCGTGTTC   | 2741 |
| Db | 1620 | TCATCTGGAAGGTCGCTCGGGCTGAGAGTGTGTTCAACAGTTAAGAACCCGTGTTC   | 1679 |
| QY | 2742 | TCTCTCCAGACACCCAGGACTTGTGAGGCTCATACAGCAACTCCAGCCCTGGGGC    | 2801 |
| Db | 1680 | TCTCTCCAGACACCCAGGACTTGTGAGGCTCATACAGCAACTCCAGCCCTGGGGC    | 1739 |
| QY | 2802 | ATCTGAGCCCTCTCTGACTTCTAAGGGCATTTGGTGACACGTACATACAGCAGGCA   | 2861 |
| Db | 1740 | ATCTGAGCCCTCTCTGACTTCTAAGGGCATTTGGTGACACGTACATACAGCAGGCA   | 1799 |
| QY | 2862 | AAACATTTATACATGTAACGTAAATGCAATAGTTAGTGAGACGCTGAAGAAAGGA    | 2921 |
| Db | 1800 | AAACATTTATACATGTAACGTAAATGCAATAGTTAGTGAGACGCTGAAGAAAGGA    | 1859 |
| QY | 2922 | GTTTATAGTCAAGGTTTAGTCTGACCTGAGTGCCCTTCCCTTAGCCGATGGTCCCT   | 2981 |
| Db | 1860 | GTTTATAGTCAAGGTTTAGTCTGACCTGAGTGCCCTTCCCTTAGCCGATGGTCCCT   | 1919 |
| QY | 2982 | AAAGTCAAGGAAAGCAATTTCTTTTATCAGGAAAGGACTTATCCCTTAGGGCTCT    | 3041 |
| Db | 1920 | AAAGTCAAGGAAAGCAATTTCTTTTATCAGGAAAGGACTTATTTCCCTTAGGGCTCT  | 1979 |
| QY | 3042 | GCTGACATCCCTAGGAACAGAGATAAATACGATGGATGTAATGAACATGCTTGGGTG  | 3101 |
| Db | 1980 | GCTGACATCCCTAGGAACAGAGATAAATACGATGGATGTAATGAACATGCTTGGGTG  | 2039 |
| QY | 3102 | AAAGGAGCCGAGTACCTGACTGGACCCAGTGGGCACTTCCACAAGCGAAGCCCGTAAC | 3161 |
| Db | 2040 | AAAGGAGCCGAGTACCTGACTGGACCCAGTGGGCACTTCCACAAGCGAAGCCCGTAAC | 2099 |
| QY | 3162 | CGATGTCACCTCAGAAATCTTCTCCCTTCTGGTAGAGATGACATCGATGATGTCGT   | 3221 |
| Db | 2100 | CGATGTCACCTCAGAAATCTTCTCCCTTCTGGTAGAGATGACATCGATGATGTCGT   | 2159 |
| QY | 3222 | GACCACTGGGCTGATGGGTGCGGCTCTGTTCTGTTCCGAATCTACCTGAGATCTC    | 3281 |
| Db | 2160 | GACCACTGGGCTGATGGGTGCGGCTCTGTTCTGTTCCGAATCTACCTGAGATCTC    | 2219 |
| QY | 3282 | AGGACAGAGGAAGCCATGAATGTTTACCAAGTGGTCTGAGTGTGATTTTACAGTTT   | 3341 |
| Db | 2220 | AGGACAGAGGAAGCCATGAATGTTTACCAAGTGGTCTGAGTGTGATTTTACAGTTT   | 2279 |
| QY | 3342 | TGAACCAATATGTTTATAGGAAATTTCTCTCTCTAGTGCCTGTGATGCAAAAGC     | 3401 |
| Db | 2280 | TGAACCAATATGTTTATAGGAAATTTCTCTCTCTAGTGCCTGTGATGCAAAAGC     | 2339 |
| QY | 3402 | CAGCCCTTCAAGTGTTCCTTCTGCTCTGCTTCCCTTGTAAATGATGCACTCGGGAA   | 3461 |
| Db | 2340 | CAGCCCTTCAAGTGTTCCTTCTGCTCTGCTTCCCTTGTAAATGATGCACTCGGGAA   | 2399 |
| QY | 3462 | ATGGGTGTCTACCCAGGGAAGGTCCTTACCCAGTCAAGGCTCAACAGTGTGTGTGA   | 3521 |
| Db | 2400 | ATGGGTGTCTACCCAGGGAAGGTCCTTACCCAGTCAAGGCTCAACAGTGTGTGTGA   | 2459 |

|    |      |   |      |
|----|------|---|------|
| QY | 3522 | GGATCAACATGCTCTGTGAAATACTGCCACCATTTCTATTGGTTGGATTTCTCAGGA   | 3581 |
| Db | 2460 | GGATCAACATGCTCTGTGAAATACTGCCACCATTTCTATTGGTTGGATTTCTCAGGA   | 2519 |
| QY | 3582 | GTCTGAATCTTCCCTCAGAGTCTCTTCTTCCCAACCCCTACCAGAGCCAACTGGGA    | 3641 |
| Db | 2520 | GTCTGAATCTTCCCTCAGAGTCTCTTCTTCCCAACCCCTACCAGAGCCAACTGGGA    | 2579 |
| QY | 3642 | TTTGAACCTTCTGACTCTCTTCTTCCCTCAGGCTGACAACTAATGGTCTCTGGGAC    | 3701 |
| Db | 2580 | TTTGAACCTTCTGACTCTCTTCTTCCCTCAGGCTGACAACTAATGGTCTCTGGGAC    | 2639 |
| QY | 3702 | ACCAGTAGGGCTTCCCAACTCTTATCCAGCTGAACTTGATTTCTTCCCAACGAG      | 3761 |
| Db | 2640 | ACCAGTAGGGCTTCCCAACTCTTATCCAGCTGAACTTGATTTCTTCCCAACGAG      | 2699 |
| QY | 3762 | GCTTGACCTGGGTCTGTTGGTCCCACTGGCCCAACACACATCTTTGGCCAGATTGGAT  | 3821 |
| Db | 2700 | GCTTGACCTGGGTCTGTTGGTCCCACTGGCCCAACACACATCTTTGGCCAGATTGGAT  | 2759 |
| QY | 3822 | TCTCAATAGATTTTATAGACATTTATCTCCACAGACTTTTAAACATGGCTTGTCTT    | 3881 |
| Db | 2760 | TCTCAATAGATTTTATAGACATTTATCTCCACAGACTTTTAAACATGGCTTGTCTT    | 2819 |
| QY | 3882 | CCATACATCCGGTCAAGTTTAAACTATTTTATACCAAGCAATTAACCAAGCAAA      | 3941 |
| Db | 2820 | CCATACATCCGGTCAAGTTTAAACTATTTTATACCAAGCAATTAACCAAGCAAA      | 2879 |
| QY | 3942 | TAGAGTACTTTCAGATATAAACTGTGTTTTCATCTTTATGTAGAGTGTCTATGTATAGG | 4001 |
| Db | 2880 | TAGAGTACTTTCAGATATAAACTGTGTTTTCATCTTTATGTAGAGTGTCTATGTATAGG | 2939 |
| QY | 4002 | CGGTATGATACCTGGCTGAAGTAAATTAACCATAGCTCTGGGAGGATTTACAGACCTT  | 4061 |
| Db | 2940 | CGGTATGATACCTGGCTGAAGTAAATTAACCATAGCTCTGGGAGGATTTACAGACCTT  | 2999 |
| QY | 4062 | TGCACTTTATGCTTTTGTGAACTCTGATAACCATGTCATATTAAGCAATTAACG      | 4121 |
| Db | 3000 | TGCACTTTATGCTTTTGTGAACTCTGATAACCATGTCATATTAAGCAATTAACG      | 3059 |
| QY | 4122 | GCATTTTCTGTG 4133   |      |
| Db | 3060 | GCATTTTCTGTG 3071   |      |

RESULT 2  
BC023423  
LOCUS Mus musculus, clone IMAGE:5010167, mRNA.  
DEFINITION BC023423  
ACCESSION BC023423  
VERSION BC023423.1 GI:19484055  
KEYWORDS HTC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Strausberg, R.  
Direct Submission  
Submitted (05-FEB-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapsb@mail.nih.gov](mailto:cgapsb@mail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,  
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,  
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 45 Row: C Column: 7  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Hexamer frequency ORF  
 analysis  
 This clone has the following problem: frame shifted.

## FEATURES

source

Location/Qualifiers

1..1341  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N-3"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5010167"  
 /tissue\_type="Mammary tumor, MMTV-LTR/INT3 model. 5 month  
 old mouse, taken by biopsy."  
 /clone\_lib="NCI CGAP\_Mam2"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"

## ORIGIN

Query Match 32.1%; Score 1341; DB 3; Length 1341;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2829 GGCATTGTCACACGTCATACAGCGAGGCAAAACATTTATACATGTAACTAATAA 2888  
 Db 1 GGCATTGTCACACGTCATACAGCGAGGCAAAACATTTATACATGTAACTAATAA 60

QY 2889 TGCATAAGTTAGTGAGACGGCTGAAGAAAGAGTTTTAGATGCAAGGTTAGTGTACC 2948  
 Db 61 TGCATAAGTTAGTGAGACGGCTGAAGAAAGAGTTTTAGATGCAAGGTTAGTGTACC 120

QY 2949 TCAGTGCCTTCTTAGCGCGATGTGTCCTCAAGTCAGGAGAACCATTTCTTTT 3008  
 Db 121 TCAGTGCCTTCTTAGCGCGATGTGTCCTCAAGTCAGGAGAACCATTTCTTTT 180

QY 3009 TATCAGGAAAGACATTTCCCTAGGGCTCTGCTGACATCCCTAGGACAGAGATAA 3068  
 Db 181 TATCAGGAAAGACATTTCCCTAGGGCTCTGCTGACATCCCTAGGACAGAGATAA 240

QY 3069 ATACGATGGATGTGAATGAACATGCTGGGTGAAGAGCGAGTACCTGACGACCC 3128  
 Db 241 ATACGATGGATGTGAATGAACATGCTGGGTGAAGAGCGAGTACCTGACGACCC 300

QY 3129 AGTGGCCACTTCCCAAGCGAGCCGGTAACCGATGTCACCTCCAGATCTTCTCCCT 3188  
 Db 301 AGTGGCCACTTCCCAAGCGAGCCGGTAACCGATGTCACCTCCAGATCTTCTCCCT 360

QY 3189 TCTGTGAGATGACATCGATGATGTGTCGACCACTGGGCCCTGATGGGTGTCGG 3248  
 Db 361 TCTGTGAGATGACATCGATGATGTGTCGACCACTGGGCCCTGATGGGTGTCGG 420

QY 3249 GCTCTGTTCTTCCGAATCTACCTGAGATCTCAGGACAGAGAGCCATGAATGTACC 3308  
 Db 421 GCTCTGTTCTTCCGAATCTACCTGAGATCTCAGGACAGAGAGCCATGAATGTACC 480

QY 3309 AAGTGGTCTGTCGTCAGTGATTTTACAGATTTTGAACCATTTGTTTTAGAGAGAT 3368  
 Db 481 AAGTGGTCTGTCGTCAGTGATTTTACAGATTTTGAACCATTTGTTTTAGAGAGAT 540

QY 3369 TCTTCTCTCTAGTGCCTGTGATGCCAAAGCCGCTTCCAGAGTGTCTCTCTGTC 3428  
 Db 541 TCTTCTCTCTAGTGCCTGTGATGCCAAAGCCGCTTCCAGAGTGTCTCTCTGTC 600

QY 3429 TGTCTCCCTTTGATGTATGATCCACTCGGGAATGGGTGTCATCCAGGAAAGTGCC 3488  
 Db 601 TGTCTCCCTTTGATGTATGATCCACTCGGGAATGGGTGTCATCCAGGAAAGTGCC 660

QY 3489 TACCCAGTCACGGGTCAACAGTGTGTGTGAGGATCAAAACATGCTCTGTGAAATAC 3548  
 Db 661 TACCCAGTCACGGGTCAACAGTGTGTGTGAGGATCAAAACATGCTCTGTGAAATAC 720

QY 3549 TGCACCCCATTTCTATTGGTTGGATTTCTCAGAGTCTGAATCTTCCCTCAGAGTCTCT 3608  
 Db 721 TGCACCCCATTTCTATTGGTTGGATTTCTCAGAGTCTGAATCTTCCCTCAGAGTCTCT 780

QY 3609 TCTCCCAACCCCTTACCCAGAGCCAAACACTGGGATTTGAACCTTCTCTGACTCTCTTCTTC 3668  
 Db 781 TCTCCCAACCCCTTACCCAGAGCCAAACACTGGGATTTGAACCTTCTCTGACTCTCTTCTTC 840

QY 3669 CTTCAAGTCTGACAACTAATGTCTCTGGGACACCCAGTGGGCTTCCCAACTCTCT 3728  
 Db 841 CTTCAAGTCTGACAACTAATGTCTCTGGGACACCCAGTGGGCTTCCCAACTCTCT 900

QY 3729 TATCAGCTGAACTTGGATTTCTCCCAACAGGGCTTGACCTGGGTGCTGTGGTCCCAC 3788  
 Db 901 TATCAGCTGAACTTGGATTTCTCCCAACAGGGCTTGACCTGGGTGCTGTGGTCCCAC 960

QY 3789 TGGCCAAACAAACATCTTTGGCCAGATGGGATTTCTCAATAGATTTTATAGACATTTATTC 3848  
 Db 961 TGGCCAAACAAACATCTTTGGCCAGATGGGATTTCTCAATAGATTTTATAGACATTTATTC 1020

QY 3849 TCCCAAGACTTTTAAACATGGTGTGTCTTCCATACATCCCGTCCAGATTTTAAAC 3908  
 Db 1021 TCCCAAGACTTTTAAACATGGTGTGTCTTCCATACATCCCGTCCAGATTTTAAAC 1080

QY 3909 TATTTTATAACCAAGGATTTAAACCAAGCAATAGAGTACTTTTCAATATAAATCTGTGT 3968  
 Db 1081 TATTTTATAACCAAGGATTTAAACCAAGCAATAGAGTACTTTTCAATATAAATCTGTGT 1140

QY 3969 TTCAATATTTATGTAGTGTCTATGTATAGCGGTATGTACCTGGTGAAGTAAATAT 4028  
 Db 1141 TTCAATATTTATGTAGTGTCTATGTATAGCGGTATGTACCTGGTGAAGTAAATAT 1200

QY 4029 TAAACATAGCTCTGGGAGATTTACAGACCTTTTGCACATTTATGCTTTTGTGAACCTCT 4088  
 Db 1201 TAAACATAGCTCTGGGAGATTTACAGACCTTTTGCACATTTATGCTTTTGTGAACCTCT 1260

QY 4089 GATAACCATGTGCAATATTTAAAGCAATTAACCTGGCATTTTCTGTAATAAATGATAT 4148  
 Db 1261 GATAACCATGTGCAATATTTAAAGCAATTAACCTGGCATTTTCTGTAATAAATGATAT 1320

QY 4149 GTATCTTAAATAAAAAAAAAAAAA 4169  
 Db 1321 GTATCTTAAATAAAAAAAAAAAAA 1341

## RESULT 3

BU511164

LOCUS

DEFINITION

IMAGE:6505924 5', mRNA sequence.

ACCESSION

BU511164

VERSION

BU511164.1

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

REFERENCE

1 (bases 1 to 1022)

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

Unpublished (1999)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Dr. David Rowe  
 cDNA Library Preparation: Invitrogen Corp  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be

BU511164 1022 bp mRNA linear EST 12-SBP-2002  
 AGENCOURT 10107530 NIH MGC 134 Mus musculus cDNA clone  
 IMAGE:6505924 5', mRNA sequence.

BU511164  
 BU511164.1 GI:22817397  
 EST.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 NIH-MGC <http://mgc.ncl.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Dr. David Rowe  
 cDNA Library Preparation: Invitrogen Corp  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM14067 row: j column: 05

High quality sequence stop: 681.

## FEATURES

source  
1..1022  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6505924"  
/tissue\_type="undifferentiated limb"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_134"  
/notes="vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;  
Cloned unidirectionally. Primer: Oligo dt. Average insert  
size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:  
this is a NIH\_MGC Library."

## ORIGIN

Query Match 18.7%; Score 780; DB 5; Length 1022;  
Best Local Similarity 96.8%; Pred. No. 7.9e-204;  
Matches 871; Conservative 0; Mismatches 20; Indels 9; Gaps 7;

```
QY 10 GCTGCTCCGATGCTCCAGAGCGGCATGGCGCCCGCCACTGGTGGGACACCTGCGGGC 69
DB 95 GCTGCTCCGATGCTCCAGAGCGGCATGGCGCCCGCCACTGGTGGGACACCTGCGGGC 154
QY 70 TGGCAGTTCGGAGTGGATTGGTGGGAGGACAACTACACTATCTGCTGCTGCATTGCCGA 129
DB 155 TGGCAGTTCGGAGTGGATTGGTGGGAGGACAACTACACTATCTGCTGCTGCATTGCCGA 214
QY 130 GTTCTACACACGATCAGCAAGCTCTGTTTTCATTTTACCTCCCATCTGCATGTCCTT 189
DB 215 GTTCTACACACGATCAGCAAGCTCTGTTTTCATTTTACCTCCCATCTGCATGTCCTT 274
QY 190 GTTCCGCCAGTACGCAAGCTCTCAACAGCGCATCTACTTATATGAGCGTCCCTAGT 249
DB 275 GTTCCGCCAGTACGCAAGCTCTCAACAGCGCATCTACTTATATGAGCGTCCCTAGT 334
QY 250 TGTAGTGGGATGGATCTGTCTACTTCCATGCAACGCTGAGTTTCCTGGGTCCAGATGCT 309
DB 335 TGTAGTGGGATGGATCTGTCTACTTCCATGCAACGCTGAGTTTCCTGGGTCCAGATGCT 394
QY 310 TGATGAATCTGCCATTCCTGCGGTTCTGATGTGCTTTTGCCCATGTTTCCAGAG 369
DB 395 TGATGAATCTGCCATTCCTGCGGTTCTGATGTGCTTTTGCCCATGTTTCCAGAG 454
QY 370 GTATTTTACCAAGATCTTTTCGAATGACAGGCGAGTTCAAGCGAGTGTGTGTCCT 429
DB 455 GTATTTTACCAAGATCTTTTCGAATGACAGGCGAGTTCAAGCGAGTGTGTGTCCT 514
QY 430 GTCTGCAATTACACGCTGCTTGGCGTTTATCAAGCCGCCATCAACAATATTTCCCTGAT 489
DB 515 GTCTGCAATTACACGCTGCTTGGCGTTTATCAAGCCGCCATCAACAATATTTCCCTGAT 574
QY 490 GATTCTGGGACTTCCATGCTGCTGCTGCTTTTGAGAGCTGAAGAGGTGTGCAATGT 549
DB 575 GATTCTGGGACTTCCATGCTGCTGCTGCTTTTGAGAGCTGAAGAGGTGTGCAATGT 634
QY 550 GCGTGTGTTTAAAGCTGGGCGCTTCTCTGCGCTCTGTTGGAAGTCTT-GGCTCTCTTCTGCT 608
DB 635 GCGTGTGTTTAAAGCTGGGCGCTTCTCTGCGCTCTGTTGGAAGTCTTGGGCTCTCTTCTGCT 694
QY 609 GGATCAGCGACCAAGCGCTTCTGTGAGCTGTCTCTCTCTTTCACCTTCCCTACCTGCACT 668
DB 695 GGATCAGCGACCAAGCGCTTCTGTGAGCTGTCTCTCTCTTTCACCTTCCCTACCTGCACT 754
QY 669 GTGTGTGGCATATTTCTCATCTGCGCTTGTGTTACCTGGGCTGTGTGTGCTTCCGCTACT 728
DB 755 GTGTGTGGCATATTTCTCATCTGCGCTTGTGTTACCTGGGCTGTGTGTGCTTCCGCTACT 814
QY 729 TTGATGTCGCTCAGAGATACCTGAGCAAGTCCAGT-CATCAGATTTCT-GGCCACGGA 786
```

```
DB 815 TTGATGTCGCTCAGAGATACCTGAGCCAAAGTCCAGTCCATCAGATTCTGGGCCACCGA 874
QY 787 GAAATGGGC--TTTATTGGTGTCCCTTATGTGTCCCTTCTGTGTGCC-ACAAAGAGTC 843
DB 875 GGAATGGCCCTTTATTGGGGGTCCCTATGGTCCCTTCTGTGTGCCAACAGAAAGTC 934
QY 844 GCCAGTCAAGAT-CACGTGATGGCAAGGAGTGA--CCAGTCTTCTACTTACTTCTTATT 900
DB 935 GCCAGTCAAGATCCACGTGATGGCAAGGAGTGAACCAAGCGCTCCCTACTTACTTCCAAT 994
```

## RESULT 4

## CN537424

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## 1

## (bases 1 to 717)

## NIH-MGC

## http://mgc.nci.nih.gov/.

## National Institutes of Health,

## Unpublished (1999)

## Contact: Robert Strausberg,

## Email: coapbs-x@mail.nih.gov

## Tissue Procurement: Dr. James Lin

## University of Iowa

## cDNA Library preparation: Dr. M. Bento Soares,

## University of Iowa

## cDNA Library Arrayed by: Dr. M. Bento Soares,

## University of Iowa

## DNA Sequencing by: Dr. M. Bento Soares,

## University of Iowa

## Clone Distribution: Distribution information can be found at

## http://genome.uiowa.edu/distribution/mousefl.html

## This clone was contributed by the Brain Molecular Anatomy Project

## (BMAP)

## Seq primer: pYX-5.

## Location/Qualifiers

## 1..717

## /organism="Mus musculus"

## /mol\_type="mRNA"

## /strain="C57BL/6"

## /db\_xref="taxon:10090"

## /clone="IMAGE:30672249"

## /tissue\_type="Upper Head"

## /dev\_stage="embryo 9.5 - 10.5 dpc"

## /lab\_host="DH10B (T1 phage resistant)"

## /clone\_lib="NIH\_BMAP\_HSO"

## /notes="Organ: Upper Head; Vector: pYX-Asc; Site 1: EcoR I;

## Site 2: Not I; The library was constructed according

## Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

## 1996. Denatured RNA was size fractionated on a 1% agarose

## gel. First strand cDNA synthesis was primed with oligo-dT

## primer containing a Not I site. Double strand cDNA was

## size selected according to mRNA size fraction, ligated

## with EcoR I adaptor, digested with NotI and then cloned

## directionally into pYX-Asc vector. The library tag

## sequence located between the Not I site and the polyA tail

## is CGACATCAAT. This library was created for the University

## Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the

## Developing Mouse Nervous System', supported by National

## Institute of Mental Health (NIMH)."

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## 715; Conservative

## 0; Mismatches

## 2; Indels

## 0; Gaps

## 0;

## QY

## 717

## GCTTCGCTACTTTGATGCTGCCTCAGAGATACCTGAGCAAGGTCCAGTTCATCAGATTCT

## 776

## DB

## 1

## GCTTCGCTACTTTGATGCTGCCTCAGAGATACCTGAGCAAGGTCCAGTTCATCAGATTCT

## 60





Db 63 CAGTGTGTTGTTGAGGATCAACATGCTCTGTGAAATATCTGCCACCCATCTTATTTGGT 122  
 QY 3568 TGGATTCTCAGGAGTCTGAATCTTCCCTCACGAGTCTCTTCTTCCCAACCCCTTACCAG 3627  
 Db 123 TGGATTCTCAGGAGTCTGAATCTTCCCTCACGAGTCTCTTCTTCCCAACCCCTTACCAG 182  
 QY 3628 AGCAACACTGGGATTTGAACCTTCTCTGACTCTCTTCTTCCCTCAGGCTGACAACTAA 3687  
 Db 183 AGCAACACTGGGATTTGAACCTTCTCTGACTCTCTTCTTCCCTCAGGCTGACAACTAA 242  
 QY 3688 TGTCTCTGGGGACACCCAGCTAGGGGCTTCCCAACTCTCTTATCCAGCTGAACCTTCGAT 3747  
 Db 243 TGTCTCTGGGGACACCCAGCTAGGGGCTTCCCAACTCTCTTATCCAGCTGAACCTTCGAT 302  
 QY 3748 TCTTCCCAACCCAGGCTTGAACCTTCTCTGACTCTCTTCTTCCCTCAGGCTGACAACTCTTT 3807  
 Db 303 TCTTCCCAACCCAGGCTTGAACCTTCTCTGACTCTCTTCTTCCCTCAGGCTGACAACTCTTT 362  
 QY 3808 GGCAGATTGGGATCTCAATAGATTTTATAGACATTTATCTCCACAGACTTTAAACA 3867  
 Db 363 GGCAGATTGGGATCTCAATAGATTTTATAGACATTTATCTCCACAGACTTTAAACA 422  
 QY 3868 TGGCTTGTGCTTTCATACATACATCCGCTCAGATTTTAAACTATTTTATTAACACAGAA 3927  
 Db 423 TGGCTTGTGCTTTCATACATACATCCGCTCAGATTTTAAACTATTTTATTAACACAGAA 482  
 QY 3928 TTAACCAACCAATAGAGTACTTTCAGATATATAACTGTGTTTCATATCTTTATGTAGAGT 3987  
 Db 483 TTAACCAACCAATAGAGTACTTTCAGATATATAACTGTGTTTCATATCTTTATGTAGAGT 542  
 QY 3988 GTGCTATGATAGCGGTATGATCCCTGGCTGAAGTAAATATTAACCATAGCTCTGGAGG 4047  
 Db 543 GTGCTATGATAGCGGTATGATCCCTGGCTGAAGTAAATATTAACCATAGCTCTGGAGG 602  
 QY 4048 ATTACAGACTTTTGCACCTTATGCTTTTGTGAACCTCTGATAACCATGTCATATTT 4107  
 Db 603 ATTACAGACTTTTGCACCTTATGCTTTTGTGAACCTCTGATAACCATGTCATATTT 662  
 QY 4108 AAAGCCAAATAGCGCAATTTTCTGTG 4133  
 Db 663 AAAGCCAAATAGCGCAATTTTCTGCG 688

RESULT 6  
 AY401891  
 LOCUS  
 DEFINITION  
 Mus musculus HCM1041 gene, VIRTUAL TRANSCRIPT, partial sequence,  
 genomic survey sequence.  
 AY401891  
 VERSION  
 AY401891.1 GI:39757877  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 1 (bases 1 to 681)  
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,  
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
 Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,  
 Adams,M.D. and Cargill,M.  
 Adams,M.D. and Cargill,M.  
 Inferring nonneutral evolution from human-chimp-mouse orthologous  
 gene trios  
 Science 302 (5652), 1960-1963 (2003)  
 JOURNAL  
 PUBLISHED  
 14671302  
 REFERENCE  
 2 (bases 1 to 681)  
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,  
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
 Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,  
 Adams,M.D. and Cargill,M.  
 Adams,M.D. and Cargill,M.  
 Direct Submission  
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD 20850, USA  
 JOURNAL  
 COMMENT  
 This sequence was made by sequencing genomic exons and ordering

them based on alignment.  
 Location/Qualifiers  
 1..681  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 <1..5681  
 /locus\_tag="HCM1041"  
 ORIGIN  
 Query Match 16.3% Score 681; DB 9; Length 681;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-176; Indels 0; Gaps 0;  
 Matches 681; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 182 ATGTGCTTGTTCGCCAGTACGCAACGCTTCAACAGCGCATCTACTTAAATATGAGC 241  
 Db 1 ATGTGCTTGTTCGCCAGTACGCAACGCTTCAACAGCGCATCTACTTAAATATGAGC 60  
 QY 242 CTCTAGTGTAGTGGGATTTGGATCTGTCTACTTCCATGCAAGCTGAGTTTCTGGGT 301  
 Db 61 CTCTAGTGTAGTGGGATTTGGATCTGTCTACTTCCATGCAAGCTGAGTTTCTGGGT 120  
 QY 302 CAGATGCTTCATGAACCTTGCCATTCTGTGGGTCTGTATGTGTCTTTGGCCATGTGTTT 361  
 Db 121 CAGATGCTTCATGAACCTTGCCATTCTGTGGGTCTGTATGTGTCTTTGGCCATGTGTTT 180  
 QY 362 CCCAGGAGGTATTTACCAAGATCTTTTCGGAATGACAGGGGCGAGTTTCAAGGCGAGTGGT 421  
 Db 181 CCCAGGAGGTATTTACCAAGATCTTTTCGGAATGACAGGGGCGAGTTTCAAGGCGAGTGGT 240  
 QY 422 TGTGTCCTGTCTGCAATTACAAAGTGTCTTGGCGTTTATCAAGCCCGGCATCAACAATTT 481  
 Db 241 TGTGTCCTGTCTGCAATTACAAAGTGTCTTGGCGTTTATCAAGCCCGGCATCAACAATTT 300  
 QY 482 TCCCTGATGATTTCTGGGACTTCCATGCACTGCGTCTGTTGTCAGAGCTGAAGAGTGT 541  
 Db 301 TCCCTGATGATTTCTGGGACTTCCATGCACTGCGTCTGTTGTCAGAGCTGAAGAGTGT 360  
 QY 542 GACAAATGCGTGTGTTTAAAGCTGGGCTCTTCTCTGGCCCTCTGTGGTGGACTCTGGCTTC 601  
 Db 361 GACAAATGCGTGTGTTTAAAGCTGGGCTCTTCTCTGGCCCTCTGTGGTGGACTCTGGCTTC 420  
 QY 602 TTCTGCTGGATCAGCAGCAAGCCCTCTGTGAGCTGTCTCTCTTCACTTCCCTTAC 661  
 Db 421 TTCTGCTGGATCAGCAGCAAGCCCTCTGTGAGCTGTCTCTCTTCACTTCCCTTAC 480  
 QY 662 CTGCACCTGTGTGGCATATTTCTCATCTGCTTCTGTTCTGTTCTGTTCTGTTCTGTTCT 721  
 Db 481 CTGCACCTGTGTGGCATATTTCTCATCTGCTTCTGTTCTGTTCTGTTCTGTTCTGTTCT 540  
 QY 722 GCCTACTTTGATGCTGCTCAGAGATACCTGAGCAAGGTCCAGTCATCAGATTCTGGCCC 781  
 Db 541 GCCTACTTTGATGCTGCTCAGAGATACCTGAGCAAGGTCCAGTCATCAGATTCTGGCCC 600  
 QY 782 AGCGAGAAATGGGCTTTTATTTGGTGTCCCTTATGTTGTCCTTCTGTTCTGTTCTGTTCT 841  
 Db 601 AGCGAGAAATGGGCTTTTATTTGGTGTCCCTTATGTTGTCCTTCTGTTCTGTTCTGTTCT 660  
 QY 842 TCGCCAGTCAAGATCAGCTGA 862  
 Db 661 TCGCCAGTCAAGATCAGCTGA 681  
 RESULT 7  
 CF110927  
 LOCUS  
 DEFINITION  
 Shultzomica04178 Rat lung airway and parenchyma cDNA libraries  
 Rattus norvegicus cDNA clone Contigs708 5', mRNA sequence.  
 CF110927  
 ACCESSION  
 VERSION  
 CF110927.1 GI:33167279  
 KEYWORDS  
 EST.  
 SOURCE  
 Rattus norvegicus (Norway rat)  
 ORGANISM  
 Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

#### REFERENCE AUTHORS

1 (bases 1 to 1388)  
Shultz, M.A., Zhang, L., Gu, Y.-Z., Baker, G.L., Fannuchi, M.V., Padua, A.M., Gurske, W.A., Morin, D., Penn, S.G., Jovanovich, S.B., Plopper, C.G. and Buckpitt, A.R.

#### TITLE

Gene expression analysis in response to lung toxicants: I.

#### JOURNAL COMMENT

Sequencing and microarray development  
Am. J. Respir. Cell Mol. Biol. 30 (3), 296-310 (2004)  
Contact: Shultz MA  
Dept. of Molecular Biosciences, School of Veterinary Medicine  
University of California, Davis  
1311 Haring Hall, One Shields Avenue, Davis, CA 95616, USA  
Tel: 530 752 0793  
Fax: 530 752 4698  
Email: mashultz@ucdavis.edu

Average Phred score is 20 or better. All poor quality data (Phred < 20) and vector/linker sequence has been removed.  
High quality sequence stop: 1388.

#### FEATURES

source

Location/Qualifiers  
1..1388  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="Contig3708"  
/sex="male"  
/tissue\_type="airway or parenchyma"  
/dev\_stage="adult"  
/clone\_lib="Rat lung lung airway and parenchyma cdna libraries"  
/note="Organ: lung; Vector: pGEM-11zf(-); Site 1: Eco RI; Site 2: Not I; mRNA was isolated from microdissected rat lung airways and parenchyma tissues."

#### ORIGIN

Query Match 16.1%; Score 670.6; DB 6; Length 1388;  
Best Local Similarity 87.2%; Pred. No. 1.7e-173;  
Matches 854; Conservative 0; Mismatches 104; Indels 21; Gaps 10;

QY 3141 CCACAGCGAAGCCCGGTAAACCGATGTCACCTCCAGAACTCTTCTCCCTTCTGTGTAGAGAT 3200  
DB 419 CAAGAAGCAGAGCAGCGCTAACTGATGTCACCTCCAGAACTCTGCTCCCTCCGCGTAGAGAT 478  
QY 3201 GACATCGATGATGATGTCGTGACCACTGGCCCTGCATGGGTGTGGGCTCTGTCTGT 3260  
DB 479 GACCTCAGTGAGTGATGTCATGGCCCACTGTCACGATGGCGCTCAGGTCTGTCTCTGT 538  
QY 3261 TCCGAATCTACCTGAGATCTCAGGACAGAGGAGCCATGAATGTTACCAAGTGGTCATGG 3320  
DB 539 TCCGAATCTGCTGGATCTCAGGACAAAGGAGCCATGAGGTACCAAGTGGTCATGT 598  
QY 3321 CTGTGAGTGAATTTAGATTTTGAACATATTTGTTTGTAGAA---GAATCTCTCTCTC 3377  
DB 599 CTGTGAGTGAATTTAGATTTTGAACATATTTGTTTGTAGAA---GAATCTCTCTCTC 658  
QY 3378 TCTAGTGGCTGTGATGTCACCAAGCCCTTCAGAAAGTCTCTCTCTCTCTCTCTCT 3437  
DB 659 TTTAGTGGCTGTGATGTCACCAAGCCCTTCAGAAAGTCTCTCTCTCTCTCTCTCT 718  
QY 3438 TTGTAATGTGATCCAC - TCGGGGAAATGGGTGTCTACCCAGGAAAGTGCCTACCCAG 3496  
DB 719 TTGTAATGTGATCCACGTGGGGAAATGGGTGTCTCTCTCTCTCTCTCTCTCTCT 778  
QY 3497 TCACGGGTCAACAGTGTGTGTGTAGGATCAACATGCGTCTGTGTAGAAATCTGTGCCACC 3556  
DB 779 TCACGGGTCAACAGCATTTGTGTGGAGGATCAACACGTGGTCTGTGTAGAAATCTGTGCCACC 838  
QY 3557 ATTCTATTGGTGGATTTCTCAGGAGTCTGAATCTTCCCTCAGGAGTCTCTCTCTCTCT 3616  
DB 839 ATTCTATTGGTGGATTTCTCAGGAGTCTGAATCTTCCCTCAGGAGTCTCTCTCTCTCT 898  
QY 3617 CCCCTACCCAGAGCCCAACTGGGATTTGAACCTTC---TCTGACTCTCTCTCTCTCTCT 3672

DB 899 CGCCTCCAGATC---CACTGGGATTGAACCTTCAAAATCCCTCTCTCTCTCTCTCT 955  
QY 3673 AGTCTGTGACAACTAATGGTCTCTGGGGACACCCAGCTAGGGCCCTTCCCCAACTCCCTT-AT 3731  
DB 956 AGTCTGTGACCACTAACGGTCTCTGGGGACACCTAATAGGGCCCTCCCCAGCTCTCTCTCT 1015  
QY 3732 CCAGCTGAACCTTGATTTCTTCCCAACAGGGCTTGACCTGGGTGCTGTGGTCCCACTGG 3791  
DB 1016 CCAGCCAAACTTGGATTTCTTCCCAACAGGGCTTGACCTGGGTGCTGTGGTCCCACTGC 1075  
QY 3792 CCAACAACAATCTTTGGCCAGATTTGGGATTTCTCAATAGATTTTATAGACATTTCTCTCC 3851  
DB 1076 CCAATGACACATCTTTGACCAAGATTTGGGATTTCTCAATAGATTTTATAGACATTTCTCTCC 1135  
QY 3852 CACAGACTTT---AAACAATGGCTTGTCTTTCCATACACATCCGGTTCAGATTTAAACT 3909  
DB 1136 CACAGACTTTTAAACAACTGGGTGTCTTTTCTTACACA---AGTCAGATTTTAAACT 1192  
QY 3910 ATTATTAACCAAGGAATTAACCAAGCAATAGAGTACTTTCAGATATAAACTGTGT 3969  
DB 1193 ATTTTATAACTCAATTAATTAACCAAGCAATAGAGTATTTTCAGATATATGC---CAAT 1250  
QY 3970 TCATACTTTATGTAGAGTGTCTATGTATAGGGGGTATGTACCTCTGCTGCTGAAGTAAATATT 4029  
DB 1251 TTATACTTTATGTAGAGTGTCTATGTATAGGGGGTATGTACCTCTGCTGCTGAAGTAAATATT 1310  
QY 4030 AACCATAGCTCTGGGAGGATTTACAGACCTTTTGCATTTATGC---TTTTTGTGAATCT 4088  
DB 1311 AGTCATTGCTCTGGG---GGATTTACAGATCTTTTGCATTTATGC---TTTTTGTGAATCT 1369  
QY 4089 GATACCATGCTCAATATT 4107  
DB 1370 GATAACCATGTTCAATATT 1388

#### RESULT 8

CNT721804

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 670)

Sharov, A.A., Piao, Y., Matoba, R., Dudekula, D.B., Qian, Y.,

VanBuren, V., Falco, G., Martin, P., Stagg, C.A., Bassey, U.C.,

Wang, Y., Carter, M.G., Hamatani, T., Alba, K., Akutsu, H., Sharova, L.,

Tanaka, T.S., Kimber, W.L., Yoshikawa, T., Jaradat, S.A., Pantano, S.,

Nagaraja, R., Boheler, K.R., Taub, D., Hodes, R.J., Longo, D.L.,

Schlesinger, D., Kellier, J., Klotz, B., Kelsoe, G., Umezawa, A.,

Vescovi, A.L., Rossant, J., Kunath, T., Hogan, B.L., Curci, A.,

D'Urso, M., Kelsie, J., Hide, W. and Ko, M.S.

Transcriptome analysis of mouse stem cells and early embryos

PLoS Biol. 1 (3), 410-419 (2003)

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@igsun.grc.nia.nih.gov

Plate: E0830 row: A column: 03

Seq primer: M13 Reverse

High quality sequence stop: 670

POLYA=No.

Location/Qualifiers

1..670

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="niaEST:E0830A03-5"  
 /db\_xref="taxon:10090"  
 /clone="NIA:E0830A03 IMAGE:30908450"  
 /tissue type="4-cell stage embryo"  
 /dev stage="4-cell"  
 /lab\_host="DH10B"  
 /clone\_lib="NIA Mouse four-cell-Embryo cDNA library (long)"  
 /notes="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library [Ref]. Genome Res. 11: 1553-1558 (2001). [PMID: 1154199]. The mRNAs were extracted from a pool of 360 embryos at 4-cell stage. Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen].  
 5'-pGAGTGTCTAGATCGGAGCGGCCCTTTT-3' from 10.8ng of mRNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker IL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2Kb. The library was constructed by Yulan Piao."

## ORIGIN

Query Match 16.0%; Score 668.4; DB 7; Length 670;  
 Best Local Similarity 99.9%; Pred. No. 5.5e-173;  
 Matches 669; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2210 ATGACTTCATCTGCTTTTGGCTTTGAGTTTAAAGTAACTGCTGGGCTATTTC 2269  
 DB 1 ATGACTTCATCTGCTTTTGGCTTTGAGTTTAAAGTAACTGCTGGGCTATTTC 60

QY 2270 TTGTTTATAGTCACAAATATTATAGAACATGAGGTGTAATAAGTCTCTTTATTA 2329  
 DB 61 TTGTTTATAGTCACAAATATTATAGAACATGAGGTGTAATAAGTCTCTTTATTA 120

QY 2330 AATTATAGCATTTACCAACTCTCCAGGTAGCAACACACAGTCATTTGATGGCGATT 2389  
 DB 121 AATTATAGCATTTACCAACTCTCCAGGTAGCAACACACAGTCATTTGATGGCGATT 180

QY 2390 TACAAGCAGGAATACCTGTTTGAATGATGTTATATGTCATTTTACTGTAGCTTAAA 2449  
 DB 181 TACAAGCAGGAATACCTGTTTGAATGATGTTATATGTCATTTTACTGTAGCTTAAA 240

QY 2450 TATGTTTAAATGACTTTGAGTGAATGTTTGTGGCAACCTAGGGTTTATGGATCAGAA 2509  
 DB 241 TATGTTTAAATGACTTTGAGTGAATGTTTGTGGCAACCTAGGGTTTATGGATCAGAA 300

QY 2510 TTGCGGCTGAGTCTCTTTGGTTTTTGGTCTAGAAAGTTTTCAGAGGGCAAAATCAAACTAG 2569  
 DB 301 TTGCGGCTGAGTCTCTTTGGTTTTTGGTCTAGAAAGTTTTCAGAGGGCAAAATCAAACTAG 360

QY 2570 CTTTGTGTTCTGATCTTACCAAGCTCGGAGCTGCCCTTATATATTTACATAAGGA 2629  
 DB 361 CTTTGTGTTCTGATCTTACCAAGCTCGGAGCTGCCCTTATATATTTACATAAGGA 420

QY 2630 CTTTATATACATAGTCTGTTAATGTCCTGAGATGACACCTAGCTGCTTATCTCGG 2689  
 DB 421 CTTTATATACATAGTCTGTTAATGTCCTGAGATGACACCTAGCTGCTTATCTCGG 480

QY 2690 AAGGGTCTCTGGGGCTGGAGGTTGGTTCAACAGTTTAAAGAACCGGTGTTCTCTCCC 2749  
 DB 481 AAGGGTCTCTGGGGCTGGAGGTTGGTTCAACAGTTTAAAGAACCGGTGTTCTCTCCC 540

QY 2750 AGACGACCCAGCGACTTGTGAGGCTCATATACAGCAACTCCAGCCCTGGGGCATCTGACG 2809  
 DB 541 AGACGACCCAGCGACTTGTGAGGCTCATATACAGCAACTCCAGCCCTGGGGCATCTGACG 600

QY 2810 CCCTCTTCTGACTTCTAAAGGGCATTTGGTGCCACACGTATACATACAGGCGAGCAAAACATTT 2869  
 DB 601 CCCTCTTCTGACTTCTAAAGGGCATTTGGTGCCACACGTATACATACAGGCGAGCAAAACATTT 660

QY 2870 ATACATGTAA 2879  
 DB 661 ATACATGTAA 670

RESULT 9  
 AK085306  
 LOCUS  
 DEFINITION  
 Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched library, clone:D63008P07 product:similar to CANCER RELATED GENE-LIVER 1 [Mus musculus], full insert sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 AK085306  
 AK085306.1 GI:26351558  
 HTC; CAP trapper.  
 Mus musculus (house mouse)  
 Mus musculus

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 99279253  
 10349636

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 20499374  
 11042159

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 11076861

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 4 The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 5 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 1173

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 6 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, I., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuura, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,

Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

# TITLE JOURNAL

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/

URL: http://fantom.gsc.riken.jp/

Location/Qualifiers

1. .1173

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM DB:D630008P07"

/db\_xref="taxon:10090"

/clone="D630008P07"

/tissue\_type="kidney"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="0 day neonate"

72. .731

/note="unnamed protein product; putative

similar to CANCER RELATED GENE-LIVER 1 [Mus musculus]

(SPTX)AA40408, evidence: FASTY, 100%ID, 77.8%length,

match=642"

/codon\_start=1

/protein\_id="BAC39416.1"

/db\_xref="GI:26315159"

/translation="MGAFHWDLHLAGSEVDWEDNYTIVPAIFNTISNVLPFI

LPPICMCFRATCSGIYLIWLLVVGISGVYFHTLSFLQMDLDEAILVLM

CALAMFPRYLPKIFRNDGRFXAVCVLSAITTLAFIKPAINISIMILGLPCTA

LIVABELKCDNRVRFKGLFSLWLTALFCWISQAFCELLSSFFPYLHCWVSADR

G"

## ORIGIN

Query Match 16.0%; Score 667.6; DB 3; Length 1173;

Best Local Similarity 98.7%; Pred. No. 1.1e-172;

Matches 673; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

|    |     |   |     |
|----|-----|---|-----|
| QY | 10  | GCTGCTCCGATGCTCCAGAGCGCCATGGCGCCCGCACTGGTGGGACCACTGGCGG | 69  |
| DB | 47  | GCTGCTCCGATGCTCCAGAGCGCCATGGCGCCCGCACTGGTGGGACCACTGGCGG | 106 |
| QY | 70  | TGGCAGTTCCGAGGTGGATTGGTGGGAGGACCACTACACTATCGTGGCTGCCA   | 129 |
| DB | 107 | TGGCAGTTCCGAGGTGGATTGGTGGGAGGACCACTACACTATCGTGGCTGCCA   | 166 |
| QY | 130 | GTTCTACACAGATCAGCAAGCTCTGTTTTCATTTTACCTCCCATCGATGCTT    | 189 |
| DB | 167 | GTTCTACACAGATCAGCAAGCTCTGTTTTCATTTTACCTCCCATCGATGCTT    | 226 |
| QY | 190 | GTTCCGCCAGTACGCAACGTCTTCAACAGCGGCATCTACTTAATATGAGCCTCT  | 249 |
| DB | 227 | GTTCCGCCAGTACGCAACGTCTTCAACAGCGGCATCTACTTAATATGAGCCTCT  | 286 |
| QY | 250 | TGTAGTGGGATGGATCTGTCTACTTCCATCCACAGCTGAGTTCTTGGGTGATG   | 309 |
| DB | 287 | TGTAGTGGGATGGATCTGTCTACTTCCATCCACAGCTGAGTTCTTGGGTGATG   | 346 |
| QY | 310 | TGATGACTTGGCATCTCTGGTCTGTGATGTCGTTGGCCATGTTGCCAGGAG     | 369 |
| DB | 347 | TGATGACTTGGCATCTCTGGTCTGTGATGTCGTTGGCCATGTTGCCAGGAG     | 406 |

|    |     |   |     |
|----|-----|---|-----|
| QY | 370 | GTATTTACCAAGATCTTTCCGAATGACAGGGGACAGGTTCAAGGAGTGTGTGTCT | 429 |
| DB | 407 | GTATTTACCAAGATCTTTCCGAATGACAGGGGACAGGTTCAAGGAGTGTGTGTCT | 466 |
| QY | 430 | GTCCTCAATACCAAGCTGCTTGGCGTTTATCAAGCCCGCATCAACAATATTTCC  | 489 |
| DB | 467 | GTCCTCAATACCAAGCTGCTTGGCGTTTATCAAGCCCGCATCAACAATATTTCC  | 526 |
| QY | 490 | GATTTCTGGGACATCTCCATGCACTGCGCTGCTTGTTCAGAGGCTGAAGAGGTGT | 549 |
| DB | 527 | GATTTCTGGGACATCTCCATGCACTGCGCTGCTTGTTCAGAGGCTGAAGAGGTGT | 586 |
| QY | 550 | GGTGTGTTTAACTGCTGGGCTCTTCTCTGCGCTCTGCTGAGACTCTCTCTGCTG  | 609 |
| DB | 587 | GGTGTGTTTAACTGCTGGGCTCTTCTCTGCGCTCTGCTGAGACTCTCTCTGCTG  | 646 |
| QY | 610 | GATCAGGACCAAGCTTCTGTGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT  | 669 |
| DB | 647 | GATCAGGACCAAGCTTCTGTGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT  | 706 |
| QY | 670 | TGTGTGGCATATTCATCTGC                                    | 691 |
| DB | 707 | TGTGTGGAGTGTGACAGGCG                                    | 728 |

## RESULT 10

AV227941

LOCUS

DEFINITION

musculus cDNA clone 443241104 3', mRNA sequence.

ACCESSION

AV227941

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Mus musculus

REFERENCE

AUTHORS

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,

Harada, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,

Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,

Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,

Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,

Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

On Nov 1, 1999 this sequence version replaced gi:6180460.

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gs.riken.jp. URL: http://genome.gsc.riken.jp/

Carninci, P., Shibata, Y., Hayashizaki, Y., Suganuma, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,

Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.

and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,

Suganuma, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a



QY 24 CCAGAGCGGCATGGGCCCCCGACCTGGTGGGACCACTGGCGGCTGGGAGTTCGGAGG 83  
Db 1 CCAGAGCGGCATGGGCCCCCGACCTGGTGGGACCACTGGCGGCTGGGAGTTCGGAGG 60  
QY 84 TGGATTGGTGGGAGGACAACTACACTATGCTGCTGCATGCGGAGTTCACAAACGA 143  
Db 61 TGGATTGGTGGGAGGACAACTACACTATGCTGCTGCATGCGGAGTTCACAAACGA 120  
QY 144 TCAGCAACGCTCTGTTTTCATTTTACCTCCCATCTGCATGCTGCTGCTGGCCAGTACG 203  
Db 121 TCAGCAACGCTCTGTTTTCATTTTACCTCCCATCTGCATGCTGCTGCTGGCCAGTACG 180  
QY 204 CAACGCTCTTCAACAGCGGCATCTACTTAATATGAGCGCTCCTAGTGTAGTGGGATTG 263  
Db 181 CAACGCTCTTCAACAGCGGCATCTACTTAATATGAGCGCTCCTAGTGTAGTGGGATTG 240  
QY 264 GATCTGCTACTTCCATGCAACGCTGAGTTTCTTGGTCAAGTCTGTGATGAATTCGCCA 323  
Db 241 GATCTGCTACTTCCATGCAACGCTGAGTTTCTTGGTCAAGTCTGTGATGAATTCGCCA 300  
QY 324 TTCTGTGGGTTCGATGCTGCTTGGCCATGCTGCTTCCAGAGGATTTTACCAAAGA 383  
Db 301 TTCTGTGGGTTCGATGCTGCTTGGCCATGCTGCTTCCAGAGGATTTTACCAAAGA 360  
QY 384 TCCTTCGGAATGACAGCGGCAGTTCAGAGCAGTGGTGTGCTGCTGCTGCAATATACAA 443  
Db 361 TCCTTCGGAATGACAGCGGCAGTTCAGAGCAGTGGTGTGCTGCTGCTGCAATATACAA 420  
QY 444 CGTGTCTGGGTTTATCAAGCCGCCATCAACATATTTCCCTGATGATTCGGGACTTC 503  
Db 421 CGTGTCTGGGTTTATCAAGCCGCCATCAACATATTTCCCTGATGATTCGGGACTTC 480  
QY 504 CATGCACTGGCTGCTTGTGACAGCTGAAGAGGTGTGACAAATGTGCGTGTGTTAAGC 563  
Db 481 CATGCACTGGCTGCTTGTGACAGCTGAAGAGGTGTGACAAATGTGCGTGTGTTAAGC 540  
QY 564 TGGGCTCTTCTCTGGCTCTGTGAGCTGTGCTCTCTTCTGCTGATGACGACCAAG 623  
Db 541 TGGGCTCTTCTCTGGCTCTGTGAGCTGTGCTCTCTTCTGCTGATGACGACCAAG 600  
QY 624 CCTTCTGTAGCTGCTCTCTCTTCTTCACTTCCCTACCTGCTGCTGCTGCTATTC 683  
Db 601 CCTTCTGTAGCTGCTCTCTCTTCTTCACTTCCCTACCTGCTGCTGCTGCTGCTG 660  
QY 684 TCATCTGC 691  
Db 661 ACAGAGGC 668

RESULT 12  
CA463294  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CA463294  
AGENCOURT 10691314 NIH\_MGC\_169 Mus musculus cDNA clone  
IMAGE:6770649 5', mRNA sequence.  
CA463294  
CA463294.1 GI:24919646  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 797)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: Dr. Jonathan Kuo, NIMH  
cDNA Library Prepared by: Michael Brownstein Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
Plate: LICM3088 row: p column: 08  
High quality sequence stop: 536.  
Location/Qualifiers  
1. 797  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6770649"  
/lab\_host="DHI0B (Tl-phase-resistant)"  
/clone\_lib="NIH\_MGC\_169"  
/notes="Organ: Testicles; Vector: pDNR-LIB; Site 1: SfiI  
(ggccattatggcc); Site 2: SfiI (ggccctcgcc); cDNA made  
by oligo-dT priming and directionally cloned. 5' and 3'  
adaptors were used in cloning as follows:  
5'-AACAGTGGTATCAACGAGTGGCCATTACGGCCGGG-3' and  
5'-ATTCTAGAGCGGCGGCACATG-dt(30)NN-3'. Full-length  
enriched library was constructed using the Clontech  
Creator SMART kit and size-selected to contain the 0.5 kb  
size fraction. Library created in the laboratory of M.  
Brownstein (NIMH, NIH). Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 15.1%; Score 631.4; DB 6; Length 797;  
Best Local Similarity 99.5%; Pred. No. 1e-162;  
Matches 654; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 10 GCTGCTCCGATGCTCAGAGCGGCATGGGCGCCGACACTGTGGGACCACTGGCGGC 69  
Db 10 GCTGCTCCGATGCTCAGAGCGGCATGGGCGCCGACACTGTGGGACCACTGGCGGC 69  
QY 70 TGGCAGTTGGAGTGGATTGGTGGGAGGACAACTACACTATCTGCTGCCATTGCCGA 129  
Db 70 TGGCAGTTGGAGTGGATTGGTGGGAGGACAACTACACTATCTGCTGCCATTGCCGA 129  
QY 130 GTTCTACAACAGCATCAGCAACGCTTGTGTTTTCATTTTACCTCCCATCTGCAATGTCCT 189  
Db 130 GTTCTACAACAGCATCAGCAACGCTTGTGTTTTCATTTTACCTCCCATCTGCAATGTCCT 189  
QY 190 GTTCCGCGAGTACGCAACGCTTCAACAGCGGCATCTACTTAATATGAGCGTCTTAGT 249  
Db 190 GTTCCGCGAGTACGCAACGCTTCAACAGCGGCATCTACTTAATATGAGCGTCTTAGT 249  
QY 250 TGTAGTGGGATTTGGATCTGTCTACTTCCATGCAACGCTGAGTTTCTGCGTCCAGATGCT 309  
Db 250 TGTAGTGGGATTTGGATCTGTCTACTTCCATGCAACGCTGAGTTTCTGCGTCCAGATGCT 309  
QY 310 TGATGAACCTTGCCATTTCTGTTGGTCTGATGTGCTTTTGGCCATGTTTCCAGGAG 369  
Db 310 TGATGAACCTTGCCATTTCTGTTGGTCTGATGTGCTTTTGGCCATGTTTCCAGGAG 369  
QY 370 GTATTTCACAAAGATCTTTCGGAATGACAGGGCGAGTTCAAGGAGTGGTGTGTCCT 429  
Db 370 GTATTTCACAAAGATCTTTCGGAATGACAGGGCGAGTTCAAGGAGTGGTGTGTCCT 429  
QY 430 GTCTGCAATTACAAAGCTGCTTGGCGTTTATCAAGCCCGCCATCAACAATATTTCCCTGAT 489  
Db 430 GTCTGCAATTACAAAGCTGCTTGGCGTTTATCAAGCCCGCCATCAACAATATTTCCCTGAT 489  
QY 490 GATTCTGGACCTTCCATGCACTGCGCTGCTTGTGTCAGAGCTGAAGAGGTGTGACAATGT 549  
Db 490 GATTCTGGACCTTCCATGCACTGCGCTGCTTGTGTCAGAGCTGAAGAGGTGTGACAATGT 549  
QY 550 GCGTGTGTTTAAAGCTGGGCTCTTCTGCGCTCTGGTGGACTGCTGCTCTCTCTGCTG 609  
Db 550 GCGTGTGTTTAAAGCTGGGCTCTTCTGCGCTCTGGTGGACTGCTGCTCTCTCTGCTG 609  
QY 610 GATCAGCGACCAAG-CCCTTCTGTGAGTCTCT-CCCTCTTTCACCTTCCCTACCTG 664  
Db 610 GATCAGCGACCAAGCCCTTCTGTGAGTCTCTCTCCCTCTTTCACCTTCCCTACCTG 666

RESULT 13





REFERENCE 1 (bases 1 to 617)  
AUTHORS Ronaldo.M.F., Lennon.G. and Soares.M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PubMed 8889548  
COMMENT Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mEST@mail.nih.gov  
Tissue Procurement: Dr. Xin-Yuan Fu, Yale University School of Medicine  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
Seq primer: M13 REVERSE.  
Location/Qualifiers  
1. .617  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-CG0p-Bfj-a-09-0-UI"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NIH BMAP Ret4 S2"  
/notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH-BMAP Ret4 S2 library is a subtracted library, ultimately derived from mouse retina tissue libraries at various stages of development. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu. The tissue for this library was contributed by Dr. Xin-Yuan Fu, Yale University School of Medicine"

FEATURES  
source  
1. .617  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-CG0p-Bfj-a-09-0-UI"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NIH BMAP Ret4 S2"  
/notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH-BMAP Ret4 S2 library is a subtracted library, ultimately derived from mouse retina tissue libraries at various stages of development. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu. The tissue for this library was contributed by Dr. Xin-Yuan Fu, Yale University School of Medicine"

ORIGIN  
Query Match 14.2%; Score 594; DB 5; Length 617;  
Best Local Similarity 99.5%; Pred. No. 2.1e-152;  
Matches 616; Conservative 0; Mismatches 1; Indels 2; Gaps 2;  
QY 1584 GTGACCTCTGTGAGACAGGCGCTCTGTGTAATCGCGGAGGCGAGCTCTGAGGCTTGACT 1643  
Db 1 GTGACCTCTGTGAGACAGGCGCTCTGTGTAATCGCGGAGGCGAGCTCTGAGGCTTGACT 60  
QY 1644 GTGGAGCGCTGTGAGATTTGGAAAGCTTTTCTATTGGAGCTTTGGTGTGCTT 1703  
Db 61 GTGGAGCGCTGTGAGATTTGGAAAGCTTTTCTATTGGAAAGCTTTTGGTGTGCTT 120  
QY 1704 ACCTCAAGATCTGACCCCGCTTCTCTGCTGTTTTAGGGGATATGTCTTCTCGAGTG 1763  
Db 121 ACCTCAAGATCTGACCCCGCTTCTCTGCTGTTTTAGGGGATATGTCTTCTCGAGTG 180  
QY 1764 AGTAGCCTCCCTAGATGTGGCCCTCGGCTTTTGTTCATPAACCTGGTGTAAC 1823  
Db 181 AGTAGCCTCCCTAGATGTGGCCCTCGGCTTTTGTTCATPAACCTGGTGTAAC 240  
QY 1824 GGTTCCTCAGACACTTCTCTCTGATGTGGCGGTCTACTACACTGATGCGTG 1883  
Db 241 GGTTCCTCAGACACTTCTCTCTGATGTGGCGGTCTACTACACTGATGCGTG 300  
QY 1884 GTGGGGATCCCAGGAGGAGGCGGAGACTCTGACATGAATGCTCTCACTGCACTCT 1943  
Db 301 GTGGGGAT-CCAGGAGGAGGCGGAGACTCTGACATGAATGCTCTCACTGCACTCT 359  
QY 1944 ACTGCTTCTACTGGGCTCTCTCAGGGTATGAGTGGGTATGGGGTATCTCAGGATGTT 2003  
Db 360 ACTGCTTCTACTGGGCTCTCTCAGGGTATGAGTGGGTATGGGGTATCTCAGGATGTT 419

QY 2004 TGTAACTCAGGACCTTCTGCTTTCTGACATTCATTCAGTGGTGGAGCTGCACTCAG 2063  
Db 420 TGTAACTCAGGACCTTCTGCTTTCTGACATTCATTCAGTGGTGGAGCTGCACTCAG 479  
QY 2064 GGGACTGGAGTTGGAAACACCGTTCTGAGTGGTCTGAGTCTGAAAGGAGCTAGTGGG 2123  
Db 480 GGGACTGGAGTTGGAAACACCGTTCTGAGTGGTCTGAGTCTGAAAGGAGCTAGTGGG 539  
QY 2124 GTTCTGGCACTTCTAGGATCTTACTCTCTGTTTAGAACCTTACAGGGTACAAAGTGGGA 2183  
Db 540 GTTCTGGCA-TTCTAGGATCTTACTCTCTGTTTAGAACCTTACAGGGTACAAAGTGGGA 598  
QY 2184 ACTGGACTTAAAGAGTTT 2202  
Db 599 ACTGGACTTAAAGAGTTT 617

RESULT 15  
AZ411514/c  
LOCUS  
DEFINITION  
AZ411514  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AZ411514 606 bp DNA linear GSS 03-OCT-2000  
IM0184B07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0184B07 R, genomic survey sequence.  
AZ411514 GI:10535527  
GSS.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,  
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von  
Niederhausern,A. and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0184 row: B column: 07  
Seq primer: CACACAGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 606.  
Location/Qualifiers  
1. .606  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0184B07"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Ti-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adsorbed DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative



of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

| ORIGIN |      | Query Match  | 13.8%           | Score 576.8;        | DB 8;     | Length 606; |
|--------|------|--|-----------------|---------------------|-----------|-------------|
|        |      | Best Local Similarity  | 99.2%           | Pred. No. 1.2e-147; |           |             |
|        |      | Matches 601;   | Conservative 0; | Mismatches 2;       | Indels 3; | Gaps 2;     |
| Qy     | 1066 | GTATCCTTTCCAGGCATACGGGCTTGCTGGGTTATGTCCATAATGACATCAACAGAGTAG | 1125            |                     |           |             |
| Db     | 606  | GTATCCTTTCCAGGCATACGGGCTTGCTGGGTTATGTCCATAATGACATCAACAGAGTAG | 547             |                     |           |             |
| Qy     | 1126 | TTCTTTGGGAGACCTAGGGCAACCAAGTTTCTTGCTGGAGAGGCTTGCTGCTTTTC     | 1185            |                     |           |             |
| Db     | 546  | TTCTTTGGGAGACCTAGGGCAACCAAGTTTCTTGCTGGAGAGGCTTGCTGCTTTTC     | 487             |                     |           |             |
| Qy     | 1186 | CACCAITTCAGACTCTCAACCCCTGATAAACAAGACCTTCTGATTTGGTGATGAAAGG   | 1245            |                     |           |             |
| Db     | 486  | CACCAITTCAGACTCTCAACCCCTGATAAACAAGACCTTCTGATTTGGTGATGAAAGG   | 427             |                     |           |             |
| Qy     | 1246 | TTCCAGAACTTTTCATTTTGGCGGAGAACTGCTTCACAAAACCAAGTGGGCAAAA      | 1305            |                     |           |             |
| Db     | 426  | TTCCAGAACTTTTCATTTTGGCGGAGAACTGCTTCACAAAACCAAGTGGGCAAAA      | 367             |                     |           |             |
| Qy     | 1306 | CACGTGTGGGGGTGTGACTAAGACGGGTGGCTTGTCACTCGCACCTCTTAGCTTTCCCA  | 1365            |                     |           |             |
| Db     | 366  | CACGTGTGGGGGTGTGACTAAGACGGGTGGCTTGTCACTCGCACCTCTTAGCTTTCCCA  | 307             |                     |           |             |
| Qy     | 1366 | GGTTCCTCCAGTGTGTGGATCCTTCCACTACCTCTGCTGAGAGATGGAGCCACGGCTT   | 1425            |                     |           |             |
| Db     | 306  | GGTTCCTCCAGTGTGTGGATCCTTCCACTACCTCTGCTGAGAGATGGAGCCACGGCTT   | 247             |                     |           |             |
| Qy     | 1426 | CAGAGGCAAAAGCTGGCAACACCCCTCTATGCCAAAAGTACACTCCTCTTTAAGCACACA | 1485            |                     |           |             |
| Db     | 246  | CAGAGGCAAAAGCTGGCAACACCCCTCTATGCCAAAAGTACACTCCTCTTTAAGCACACA | 187             |                     |           |             |
| Qy     | 1486 | TTAC--ATAGACACTATTCTGCTCTTCCAGAGTGCAGACGCTTCAGACCCACAGAGAAAT | 1543            |                     |           |             |
| Db     | 186  | TTACATATAGACACTATTCTGCTCTTCCAGAGTGCAGACGCTTCAGACCCACAGAGAAAT | 127             |                     |           |             |
| Qy     | 1544 | CCTTCAGGTTATGGAAGATTCACACACAGCCCCCTCTTGTGACCTCTGTGAGAGCAAG   | 1603            |                     |           |             |
| Db     | 126  | CCTTCAGGTTATGGAAGATTCACACACAGCCCCCTCTTGTGACCTCTGTGAGAGCAAG   | 67              |                     |           |             |
| Qy     | 1604 | GCCTCGTTGTAATCGCGAGGACGCTCTGGAGGCTTGACTGTGGAGCGCCTGTGAAGATT  | 1663            |                     |           |             |
| Db     | 66   | GCCTCGTTGTAATCGCGAGGACGCTCTGGAGGCTTGACTGTGGAGCG-CTGTGAAGATT  | 8               |                     |           |             |
| Qy     | 1664 | TTGAAA 1669  |                 |                     |           |             |
| Db     | 7    | TGAAA 2  |                 |                     |           |             |

Search completed: November 21, 2004, 06:57:29  
Job time : 11526.4 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 21, 2004, 03:07:33 ; Search time 3473 Seconds

(without alignments)

3744.505 Million cell updates/sec

Title: US-10-017-410-2

Perfect score: 1512

Sequence: 1 MGAPHWMDHLRAGSEVDWC.....IGVYVSVLLCAHKKSPVKIT 275

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1p  
-Q/cgm2 1/USPTO.spool.p/US10017410/runat.19112004.133620.18996/app.query.fasta\_1.455  
-DB-GenEmbl -OFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPT=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10017410 @CGN 1 1 5600 @runat.19112004.133620.18996 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_bg.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | ID          | Description |
|------------|--------|-------------|--------|-------------|-------------|
| 1          | 1512   | 100.0       | 4174   | 10 AF282864 | Mus muscu   |
| 2          | 1426   | 94.3        | 2852   | 6 CQ842051  | Sequence    |
| 3          | 1426   | 94.3        | 2852   | 9 AK123581  | Homo sapi   |
| 4          | 1373.5 | 90.8        | 822    | 6 AX206799  | Sequence    |

|   |    |        |      |        |    |          |           |           |
|---|----|--------|------|--------|----|----------|-----------|-----------|
| C | 5  | 1361.5 | 90.0 | 149462 | 9  | AC017081 | AC017081  | Homo sapi |
|   | 6  | 1252   | 82.8 | 4078   | 10 | BC059819 | Mus muscu |           |
|   | 7  | 800.5  | 52.9 | 1429   | 3  | AK116177 | Ciona int |           |
|   | 8  | 714.5  | 47.3 | 2318   | 3  | AY071232 | Drosophil |           |
|   | 9  | 706.5  | 46.7 | 1163   | 3  | AF323976 | Drosophil |           |
|   | 10 | 549.5  | 36.3 | 792    | 6  | AX206797 | Sequence  |           |
|   | 11 | 549.5  | 36.3 | 795    | 9  | AF347024 | Homo sapi |           |
|   | 12 | 547    | 36.2 | 315    | 6  | CQ744096 | Sequence  |           |
|   | 13 | 545    | 36.0 | 315    | 6  | CQ736905 | Sequence  |           |
|   | 14 | 545    | 36.0 | 315    | 6  | CQ736906 | Sequence  |           |
|   | 15 | 517.5  | 34.2 | 822    | 10 | AF347023 | Mus muscu |           |
|   | 16 | 455    | 30.1 | 537    | 6  | CQ605547 | Sequence  |           |
|   | 17 | 340.5  | 22.5 | 90370  | 10 | BX005084 | Mouse DNA |           |
|   | 18 | 340.5  | 22.5 | 207360 | 2  | EX324222 | Mus muscu |           |
|   | 19 | 336.5  | 22.3 | 112713 | 9  | AL391834 | Human DNA |           |
|   | 20 | 336.5  | 22.3 | 219679 | 2  | AC109451 | Homo sapi |           |
|   | 21 | 334.5  | 22.1 | 184163 | 2  | AC120245 | Rattus no |           |
|   | 22 | 334.5  | 22.1 | 249734 | 2  | AC097362 | Rattus no |           |
|   | 23 | 320    | 21.2 | 4758   | 6  | CQ605546 | Sequence  |           |
| C | 24 | 320    | 21.2 | 13715  | 2  | AC017332 | Drosophil |           |
|   | 25 | 320    | 21.2 | 78857  | 3  | AC004364 | Drosophil |           |
|   | 26 | 320    | 21.2 | 157875 | 3  | AE003665 | Drosophil |           |
|   | 27 | 320    | 21.2 | 164361 | 3  | AC009252 | Drosophil |           |
|   | 28 | 320    | 21.2 | 165585 | 3  | AC093044 | Drosophil |           |
|   | 29 | 316.5  | 20.9 | 163542 | 9  | AL158206 | Human DNA |           |
| C | 30 | 316.5  | 20.9 | 219679 | 2  | AC109451 | Homo sapi |           |
|   | 31 | 264.5  | 17.5 | 4071   | 6  | CQ615131 | Sequence  |           |
| C | 32 | 264.5  | 17.5 | 38138  | 2  | AC017406 | Drosophil |           |
|   | 33 | 227.5  | 15.0 | 677    | 5  | CR523187 | Gallus ga |           |
|   | 34 | 219    | 14.5 | 257    | 6  | CQ731034 | Sequence  |           |
|   | 35 | 212.5  | 14.1 | 1313   | 8  | AK103953 | Oryza sat |           |
|   | 36 | 212.5  | 14.1 | 1335   | 8  | AK073303 | Oryza sat |           |
|   | 37 | 209.5  | 13.9 | 801    | 6  | AX206798 | Sequence  |           |
|   | 38 | 208.5  | 13.8 | 1347   | 9  | BC073853 | Homo sapi |           |
|   | 39 | 208.5  | 13.8 | 3404   | 9  | AF214454 | Homo sapi |           |
|   | 40 | 208.5  | 13.8 | 3473   | 9  | AF327353 | Homo sapi |           |
| C | 41 | 204    | 13.5 | 358    | 6  | AR419744 | Sequence  |           |
|   | 42 | 204    | 13.5 | 358    | 6  | AX980438 | Sequence  |           |
| C | 43 | 204    | 13.5 | 358    | 6  | BD115297 | EST and e |           |
|   | 44 | 204    | 13.5 | 799    | 8  | BT008549 | Arabidops |           |
|   | 45 | 204    | 13.5 | 1193   | 8  | AY090947 | Arabidops |           |

#### ALIGNMENTS

|            |  |  |         |      |        |                 |
|------------|--|--|---------|------|--------|-----------------|
| RESULT 1   | AF282864   | Mus musculus cancer related gene-liver 1 mRNA, complete cds. | 4174 bp | mRNA | linear | ROD 12-DEC-2001 |
| LOCUS      | AF282864   | Mus musculus cancer related gene-liver 1 mRNA, complete cds. | 4174 bp | mRNA | linear | ROD 12-DEC-2001 |
| DEFINITION | AF282864   | Mus musculus cancer related gene-liver 1 mRNA, complete cds. | 4174 bp | mRNA | linear | ROD 12-DEC-2001 |
| ACCESSION  | AF282864   | Mus musculus cancer related gene-liver 1 mRNA, complete cds. | 4174 bp | mRNA | linear | ROD 12-DEC-2001 |
| VERSION    | AF282864.1   | GI:17529683  | 4174 bp | mRNA | linear | ROD 12-DEC-2001 |
| KEYWORDS   | Mus musculus (house mouse)   |  |         |      |        |                 |
| SOURCE     | Mus musculus (house mouse)   |  |         |      |        |                 |
| ORGANISM   | Mus musculus   |  |         |      |        |                 |
| REFERENCE  | 1  | (bases 1 to 4174)  |         |      |        |                 |
| AUTHORS    | Graveel, C.R., Jatke, T., Madore, S.J., Holt, A.L., and Farnham, P.J.  |  |         |      |        |                 |
| TITLE      | Expression profiling and identification of novel genes in hepatocellular carcinomas  |  |         |      |        |                 |
| JOURNAL    | Oncogene 20 (21), 2704-2712 (2001)   |  |         |      |        |                 |
| MEDLINE    | 21313787   |  |         |      |        |                 |
| PUBMED     | 11420682   |  |         |      |        |                 |
| REFERENCE  | 2  | (bases 1 to 4174)  |         |      |        |                 |
| AUTHORS    | Graveel, C.R., Jatke, T., Madore, S.J., Holt, A.L., and Farnham, P.J.  |  |         |      |        |                 |
| TITLE      | Identification of genes deregulated in murine hepatocellular carcinomas using oligonucleotide microarrays and representational difference analysis |  |         |      |        |                 |
| JOURNAL    | Unpublished  |  |         |      |        |                 |
| REFERENCE  | 3  | (bases 1 to 4174)  |         |      |        |                 |
| AUTHORS    | Farnham, P.J. and Graveel, C.R.  |  |         |      |        |                 |
| TITLE      | Direct Submission  |  |         |      |        |                 |
| JOURNAL    | Submitted (27-JUN-2000) Oncology, University of Wisconsin, 1400  |  |         |      |        |                 |

University Avenue, Madison, WI 53706, USA

FEATURES  
 source  
 1. .4174  
 /location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /tissue\_type="hepatocellular carcinoma"  
 35. .862  
 /notes="CRG-11; putative"  
 /codon\_start=1  
 /product="cancer related gene-liver 1"  
 /protein\_id="AAL40408.1"  
 /db\_xref="GI:17529684"  
 /translation="MGAPHWDLRAGSEVDWCEENYIVPAIAFYNTISNLFPI  
 LPICMLERQVATCFNSGYLIWLLVVVGISVFHATLSFLGMDLALLWVIM  
 LLAAMFFPRYIPKIFRNDGRFKAVCVLSAITTCLAFIKPAINNISMLILGLPCTA  
 LVAELKRCNDVRFKLFGLFWTLALFCWISDQAFCELLSSFPFLYHCVWHLI  
 CLASYLGCVCFAFYDAASBIPEQGPVIRFPSEKNAFIGVPVYVSLCAHKKSPVKIT"

## ORIGIN

Alignment Scores:  
 Pred. No.: 6.94e-134 Length: 4174  
 Score: 1512.00 Matches: 275  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0

US-10-017-410-2 (1-275) x AF282864 (1-4174)

QY 1 MetGlyAlaProHisTriPAspHisLeuArgAlaGlySerSerGluValAspTrpCys 20  
 Db 35 ATGGGGCCCCGACCTGGTGGGACCACTTGGGGCTGGGAGTTCGGAGGTGGATTGGTGC 94

QY 21 GluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrAsnThrIleSerAsnVal 40  
 Db 95 GAGGACAACTACACTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 154

QY 41 LeuPhePheIleLeuProIleCysMetCysLeuPheArgGlnTyrAlaThrCysPhe 60  
 Db 155 TTGTTTTTTCATTTTACCTCCCATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 214

QY 61 AsnSerGlyIleTyrLeuIleTrpThrLeuLeuValValGlyIleGlySerValTyr 80  
 Db 215 AACAGGGGACATCTACTATATAGGAGCTTCTAGTTGTAGTGGGATTTGGAICTGCTCTAC 274

QY 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTrpVal 100  
 Db 275 TTCCATGCAACGCTGAGTTTCTGGGTGAGATGCTTGTATGATGAACTTGCCTTCTGTGGGTT 334

QY 101 LeuMetCysAlaLeuAlaMetTrpPheProArgArgTyrLeuProIlyIlePheArgAsn 120  
 Db 335 CTGATGTGCTTTGGCCCATGTGTTTCCAGAGGATATTTACCAAGATCTTTTCGGAAAT 394

QY 121 AspArgGlyArgPheIlyAlaValValCysValLeuSerAlaIleThrThrCysLeuAla 140  
 Db 395 GACAGGGGAGGTTCAGGACAGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 454

QY 141 PheIleLysProAlaIleAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla 160  
 Db 455 TTTATCAAGCCGCCCATCAACAATATTTCCCTGATGATTTCTGGACTTCCATGACATGCG 514

QY 161 LeuLeuValAlaGluLeuIlyArgCysAspAsnValArgValPheIlySerGlyLeuPhe 180  
 Db 515 CTGCTTGTTCGAGCTGAGAGGTGTGACATGTGCGTGTGTAAAGCTGGGCCCTCTTC 574

QY 181 SerGlyLeuTrpTrpThrLeuAlaLeuPheCysTrpIleSerAspGlnAlaPheCysGlu 200  
 Db 575 TCTGGCCTCTGTGGACTCTGGCTCTCTTCTGCTGGATCAGCGACCAAGCCCTCTCTGTGAG 634

QY 201 LeuLeuSerSerPheHisPheProTyrLeuHisCysValTrpHisIleLeuIleCysLeu 220  
 Db 635 CTGCTCTCTCCTTTTCACTTCCCTACCTGCATGTGTGTGGCATATTTCTCATCTGCTT 694

QY 221 AlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAlaAlaSerGluIleProGlu 240  
 Db 695 GCTTCGTACCTGGCTGTGTGCTTTCGCTACTTTTGTGCTGCTCCTCAGAGATACCTGAG 754

QY 241 GlnGlyProValIleArgPheTrpProSerGluLeuTrpAlaPheIleGlyValProTyr 260  
 Db 755 CAAGGTCCAGTATCATGATTTCTGGCCACGAGAAATGGCTTTTATTTGTGTGCTTAT 814

QY 261 ValSerLeuLeuCysAlaHisIlySerProValIlyIleThr 275  
 Db 815 GTGTCCCTTCTGTGTGCCACAAAGAGTGCCTCAGTCAAGATCAG 859

RESULT 2  
 CQ842051  
 LOCUS  
 DEFINITION  
 ACCESSION  
 CQ842051  
 VERSION  
 CQ842051.1  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 AUTHORS  
 Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,  
 Yamamoto,J., Isono,Y., Nagai,K. and Irie,R.  
 Full-length human cdna  
 Patent: EP 1440981-A 698 28-JUL-2004;  
 Research Association for Biotechnology (JP)  
 JOURNAL  
 LOCATION/Qualifiers  
 1. .2852  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 6.93e-126 Length: 2852  
 Score: 1426.00 Matches: 255  
 Percent Similarity: 97.82% Conservative: 14  
 Best Local Similarity: 92.73% Mismatches: 6  
 Query Match: 94.31% Indels: 0  
 DB: 6 Gaps: 0

US-10-017-410-2 (1-275) x CQ842051 (1-2852)

QY 1 MetGlyAlaProHisTriPAspHisLeuArgAlaGlySerSerGluValAspTrpCys 20  
 Db 159 ATGGGGCCCCGACCTGGTGGGACCACTGCGAGGTGGTAGTCTCGAGGTGGACTGGTGC 218

QY 21 GluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrAsnThrIleSerAsnVal 40  
 Db 219 GAGGACAACTACACCATCGTCTGCTATCGCGAGTCTTACACACGATCAGCATGTC 278

QY 41 LeuPhePheIleLeuProIleCysMetCysLeuPheArgGlnTyrAlaThrCysPhe 60  
 Db 279 TTATTTTTCATTTTACCGCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 338

QY 61 AsnSerGlyIleTyrLeuIleTrpThrLeuLeuValValGlyIleGlySerValTyr 80  
 Db 339 AACAGTGGCATCTACTTAACTTGGACTCTTTTGTGTAGTGGAAATGGATCCGCTAC 398

QY 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTrpVal 100  
 Db 399 TTCCATGCAACCTTAGTTCCTGGTTCAGATGCTTGTATGAACTTGCAGTCTCTTGGGTT 458

QY 101 LeuMetCysAlaLeuAlaMetTrpPheProArgArgTyrLeuProIlyIlePheArgAsn 120  
 Db 459 CTGATGTGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 518

QY 121 AspArgGlyArgPheIlyAlaValValCysValLeuSerAlaIleThrThrCysLeuAla 140



```

ACCESSION   AX206799
VERSION     AX206799.1  GI:15394643
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS     Hofmann,K. and Conradt,M.
TITLE       Ceramidase
JOURNAL     Patent: WO 0155408-A 6 02-AUG-2001;
            Memorec Medical Research Cologne Stoffel GmbH (DE)
FEATURES
    source
        1..822
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 1..63e-121 Length: 822
Score: 1373.50 Matches: 246
Percent Similarity: 94.91% Conservative: 15
Best Local Similarity: 89.45% Mismatches: 13
Query Match: 90.84% Indels: 1
DB: 6 Gaps: 1
US-10-017-410-2 (1-275) x AX206799 (1-822)

QY 1 MetGlyAlaProHisTrpTrpAspHisLeuArgAlaGlySerSerGluValAspTrpCys 20
Db 1 ATGGGGCCCGCCACTGGTGGGACCGAGCTGCAGCTGGCAGCTCGGAGGTGGACTGGCGC 60
QY 21 GluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrZnThrIleSerAsnVal 40
Db 61 GAGGACAACCTACACCATCGCTGTCTGCTGCCGAGTTCTATAACATGATCAGCAATGTC 120
QY 41 LeuPhePheIleLeuProPheCysMetCysLeuPheArgGlnTyrAlaThrCysPhe 60
Db 121 TTATTTTTCATTACCGCCCATCTGCAATGTGTTGTTCGTCAGTATGCAATGCTTC 180
QY 61 AsnSerGlyIleTyrLeuIleTrpThrLeuLeuValValGlyIleGlySerValTyr 80
Db 181 AACAGCGGCATCTACTTAATCTGG--CTCTGTTGTAGCGGGAATGGATCGGCTAC 237
QY 81 PheHisAlaThrIleuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTrpVal 100
Db 238 TTCATGCAACCTTAGTTTCCTGGGTGACATGCTTGATGAACCTTCAGTCTCTTGGGTT 297
QY 101 LeuMetCysAlaLeuAlaMetTrpPheProArgArgTyrLeuProLysIlePheArgAsn 120
Db 298 CTGATGTGTGCTTCGGTTCATGTGGTTCCTCCAGAGGATATCTACCAAGATCTTTCGGAT 357
QY 121 AspArgGlyArgPheLysAlaValValCysValLeuSerAlaIleThrCysLeuAla 140
Db 358 GACAGGGTAGGTTCAGGTGGTGGTGTGTCTGTCTGTGACGTATGACGTGGCGCA 417
QY 141 PheIleLysProAlaIleAlaSerLeuMetIleLeuGlyLeuProCysThrAla 160
Db 418 TTGTCAAGCTCCCATCAACATCTCTCTGATGACCTGGAGTTCCTTCGGTGCA 477
QY 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180
Db 478 CTGCTCATCAGAGCTAAGAGGTGTGACCAATGGTGTGTAAAGCTGGGCTCTTC 537
QY 181 SerGlyLeuTrpTrpThrLeuAlaLeuPheCysTyrTrpIleSerAspGlnAlaPheCysGlu 200
Db 538 TCGGGCTCTCTGGTGGACCTGGCCCTGTTCTGTGGATCAGTACCGAGCTTCTCGGAG 597
QY 201 LeuLeuSerSerPheHisPheProTyrLeuHisCysValTrpHisIleLeuIleCysLeu 220
Db 598 CTGCTGTATCTTCACTTCCCTACCTGCTGATGTCATGTGGCAGCATCTCTCATCTGCTT 657

```

221 AlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAlaAlaSerGluIleProGlu 240  
 658 GCTGCTACTCTGGCTGTGTATGCTTTCCTTCTGATGCTGCTCCTCAGAGATCTCTGAG 717

241 GlnGlyProValIleArgPheTrpProSerGluLysTrpAlaPheIleGlyValProTyr 260  
 718 CRAAGGCCCTGTCAATCAATTCGCCCGAGGAGAAATGGCCCTTCATTTGGTGTGCTCCCTAT 777

261 ValSerLeuLeuCysAlaHisLysLysSerProValLysIleThr 275  
 778 GTGTCCCTCTGTGTGTCACCAAGAAATCATCAGTCAAGACCAG 822

AC017081 149462 bp DNA linear PRI 09-JAN-2002  
 Homo sapiens BAC clone RP11-470J24 from 2, complete sequence.  
 AC017081  
 AC017081.8 GI:18093316  
 HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 149462)  
 AUTHORS Sulston,J.E. and Waterston,R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 PUBMED 9847074  
 REFERENCE 2 (bases 1 to 149462)  
 AUTHORS Nguyen,C., Doeber,A. and Kozlowicz,A.  
 TITLE The sequence of Homo sapiens BAC clone RP11-470J24  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 149462)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-DEC-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE 4 (bases 1 to 149462)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Jan 9, 2002 this sequence version replaced gi:14165368.

Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc  
 Contact: sapiens@wustl.edu  
 Summary Statistics  
 Center project name: H\_NH0470J24

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by  
 restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D.  
 McPherson, Department of Genetics, Washington University, St. Louis  
 MO. For additional information about the map position of this  
 sequence, see http://genome.wustl.edu/gsc

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-310K15, 2000 bp overlap; the clone sequenced to the right is RP11-156A1, 2000 bp overlap. Actual start of this clone is at base position 190775 of RP11-310K15; actual end is at base position 28935 of RP11-156A1.

The sequence between 81601 and 81703 is covered only by a PCR product of clone DNA. Unresolved tandem repeat regions exist between 81548 and 8183, 126196 and 127426.

## FEATURES

source

1. .149462

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="RP11-470J24"

/clone\_lib="RPCI-11"

141. .282

/rpt\_family="MIR"

581. .621

/rpt\_family="TGGG)n"

965. .1037

/rpt\_family="GC\_rich"

1127. .1561

/note="match to EST AA056210 (NID:g1548612) zf62g06.sl"

1771. .2127

/note="match to EST AW614820 (NID:g7320006) hg80c08.xl"

2055. .2195

/rpt\_family="MER1\_type"

2216. .2419

/rpt\_family="MER1\_type"

2604. .2844

/rpt\_family="MIR"

2991. .3324

/note="match to EST R35058 (NID:g791959) yhs6d09.r1"

3207. .3758

/note="match to EST AW452004 (NID:g6992780)"

3486. .3527

/rpt\_family="AT\_rich"

3501. .3745

/note="match to EST AA054208 (NID:g1545131) zf51g03.r1"

3981. .4017

/rpt\_family="T-rich"

4231. .4273

/rpt\_family="AT\_rich"

5518. .5847

/note="match to EST R80994 (NID:g857275) y194c08.r1"

5606. .5633

/rpt\_family="AT\_rich"

5722. .5758

/rpt\_family="AT\_rich"

5802. .5831

/rpt\_family="(TTTTA)n"

5808. .6116

/rpt\_family="Alu"

6153. .6244

/rpt\_family="MER1\_type"

6965. .7000

/rpt\_family="AT\_rich"

7020. .7094

/rpt\_family="MER1\_type"

7982. .8006

repeat\_region  
8000. .8436  
/rpt\_family="AT\_rich"  
repeat\_region  
8437. .8540  
/rpt\_family="L2"  
repeat\_region  
8541. .8696  
/rpt\_family="L1"  
repeat\_region  
8722. .9684  
/rpt\_family="L2"  
repeat\_region  
9378. .9422  
/rpt\_family="L1"  
repeat\_region  
9685. .9973  
/rpt\_family="(TTG)n"  
repeat\_region  
9974. .10197  
/rpt\_family="Alu"  
repeat\_region  
10280. .10305  
/rpt\_family="L1"  
repeat\_region  
10320. .10341  
/rpt\_family="(TATG)n"  
repeat\_region  
10333. .10679  
/rpt\_family="AT\_rich"  
repeat\_region  
10681. .10757  
/rpt\_family="L1"  
repeat\_region  
10758. .11039  
/rpt\_family="Alu"  
repeat\_region  
11040. .12011  
/rpt\_family="L1"  
repeat\_region  
11999. .12591  
/rpt\_family="L1"  
repeat\_region  
12592. .12706  
/rpt\_family="L1"  
repeat\_region  
12707. .12952  
/rpt\_family="Alu"  
repeat\_region  
12966. .13140  
/rpt\_family="L1"  
repeat\_region  
13139. .16061  
/rpt\_family="L1"  
repeat\_region  
16063. .17107  
/rpt\_family="L1"  
repeat\_region  
17108. .17161  
/rpt\_family="MaLR"  
repeat\_region  
17163. .17468  
/rpt\_family="MER1\_type"  
repeat\_region  
17471. .17675  
/rpt\_family="MaLR"  
repeat\_region  
17677. .17835  
/rpt\_family="L1"  
repeat\_region  
17909. .18144  
/rpt\_family="L1"  
repeat\_region  
18177. .18501  
/rpt\_family="L1"  
repeat\_region  
18519. .20393  
/rpt\_family="L1"  
repeat\_region  
20490. .20906  
/rpt\_family="L1"  
repeat\_region  
21258. .21365  
/rpt\_family="MIR"  
repeat\_region  
21353. .21393  
/rpt\_family="(T)n"  
repeat\_region  
21366. .21684  
/rpt\_family="Alu"  
repeat\_region  
21543. .21566  
/rpt\_family="AT\_rich"  
repeat\_region  
21685. .21819  
/rpt\_family="MIR"  
repeat\_region  
22160. .22640  
/rpt\_family="MER2\_type"  
repeat\_region  
22878. .23228

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

9.78e-118

1361.50

94.91%

Length:

Matches:

Conservative: 15

```

Best Local Similarity: 89.45% Mismatches: 14
Query Match: 90.05% Indels: 1
DB: 9 Gaps: 0

US-10-017-410-2 (1-275) x AC017081 (1-149462)

QY 1 MetGlyAlaProHisTrrTrrPheHisLeuArgAlaGlySerSerGluValAspTrrCys 20
DB 136921 ATGGGCGCCCGGACTGCTGGGACACAGCTGCGAGCTCGAGCTGGAGTGGATGGCGC 136862
QY 21 GluAspAsnTrrThrIleValProAlaIleAlaGluPheTrrAsnThrIleSerAsnVal 40
DB 136861 GAGGACAACTACACCATCGTCTGCTGCGCCGAGTCTTATAACATGATCAGCAATGTC 136802
QY 41 LeuPhePheIleLeuProProlIleCysMetCysLeuPheArgGlnTrrAlaThrCysPhe 60
DB 136801 TTAATTTTCATTTACCGCCCATCTGCATGTGCTTGTTCGTGAGTATGCAACATGCTTC 136742
QY 61 AsnSerGlyIleTrrLeuIleTrrTrrLeuValValGlyIleGlySerValTrr 80
DB 136741 AACAGCGCATCTACTTAATCTGACTC--TTGGTTGTAGCGGAAATGGATCCGCTTAC 136684
QY 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTrrVal 100
DB 136683 TTCCATGCAACCCCTTAGTTTCTGGTTCAGATGCTTGTGAACTTGCAGTCTCTTGGGTT 136624
QY 101 LeuMetCysAlaLeuAlaMetTrrPheProArgArgTrrLeuProLysIlePheArgAsn 120
DB 136623 CTGATGTGTCTCGGTCATGTGCTTCCCGAAGAGTATCTACCAAGATCTTTTCGAAT 136564
QY 121 AspArgGlyArgPheLeuAlaValValCysValLeuSerAlaIleTrrThrCysLeuAla 140
DB 136563 GACCAGGATAGTTCAGAGTGGTGTCTGTGCTCTGCTGAGTATGACGTGCTGGCA 136504
QY 141 PheIleLysProAlaIleAsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla 160
DB 136503 TTTGTCAAGCTGTCATCAACCAACATCTCTGTATGACCCCTGGGAGTCTCTTGGCTGCA 136444
QY 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180
DB 136443 CTGCTCATCATCAGAGCTAAAGAGTGTGACAAATCGTGTGTATAGCTGGGCTCTTC 136384
QY 181 SerGlyLeuTrrTrrTrrLeuAlaLeuPheCysTrrIleSerAspGlnAlaPheCysGlu 200
DB 136383 TCGGCTCTGCTGGGACCCCTGGCCCTCTCTGCTGGATCATGACCGAGCTTCTGCGAG 136324
QY 201 LeuLeuSerSerPheHisPheProTrrLeuHisCysValTrrPheIleLeuIleCysLeu 220
DB 136323 CTGCTGTATCTTCAACTTCCCTACCTGCATGTCATGTGGCACATCTCATCTGCTT 136264
QY 221 AlaSerTrrLeuGlyCysValCysPheAlaTrrPheAlaAlaSerGluIleProGlu 240
DB 136263 GCTGCCCTACCTGGGCTGTGTATGCTTGGCTTACTTGTATGCTGCTCAGAGATTCCTGAG 136204
QY 241 GlnGlyProValIleArgPheTrrProSerGluLysTrrAlaPheIleGlyValProTrr 260
DB 136203 CAAGGCCCTGTTCATCAATTCGSCCCCGAGAGAAATGGGCTTCATTTGGTGGCTTAT 136144
QY 261 ValSerLeuLeuCysAlaHisLysLeuSerProValLysIleThr 275
DB 136143 GTGTCCCTCTGTGTGTCACCAAGAAATCATCATCAAGACCAGC 136099

RESULT 6
LOCUS BC059819 4078 bp mRNA linear ROD 21-OCT-2003
DEFINITION Mus musculus cancer related gene-liver 1, mRNA (cDNA clone
MGC:69583 IMAGE:6839525), complete cds.
ACCESSION BC059819
VERSION BC059819.1 GI:37590519
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4078)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Armanaci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Carrancio, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Vallalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 4078)
Strausberg, R.
Direct Submission
Submitted (07-OCT-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgbs-x@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 132 Row: 0 Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21314857.
FEATURES
Location/Qualifiers
1..4078
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:69583 IMAGE:6839525"
/tissue_type="Brain"
/clone_lib="NIH BMAP_GH0"
/lab_host="DH10B"
/notes="Vector: pYX-ASC"
1..4078
/gene="CRG-L1"
/db_xref="LocusID:230379"
76..765
/gene="CRG-L1"
/codon_start=1
/product="CRG-L1 protein"

```









Score: 549.50 Matches: 103  
Percent Similarity: 59.38% Conservative: 49  
Best Local Similarity: 40.23% Mismatches: 102  
Query Match: 36.34% Indels: 2  
DB: 6 Gaps: 1

US-10-017-410-2 (1-275) x AX206797 (1-792)

QY 9 HisLeuArgAlaGly-SerSerGluValAspTyrCysGluAspAsnTyrThrIleValPr 28  
Db 9 CATCTTCGCTATCAGAGCTCCGAGGTGGACTGGTGTGAGAGCAACTCCAGTACTCGGA 68  
QY 28 oAlaIleAlaGluPheThrIleSerAsnValLeuPhePheIleLeuProProl 48  
Db 69 GCTGGTGGCGAGTCTTACACAGCTTCTCCAAATATCCCTTCTTATCTTCGCGGCAC 128  
QY 48 eCysMetCysLeuPheArgGlnTyrAlaThrCysPheAsnSerGlyIleTyrLeuIleTr 68  
Db 129 GATGATGCTCTGATGCACCCGATGATGCCAGAGCGCTCCCGCTACATTACGTTGTCIG 188  
QY 68 pThrLeuLeuValValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLe 88  
Db 189 GGTCTCTTCATGATCATAGGCTGTCTCCATGATATTCACATGATCCCTTCTTATCTTCGCGGCAC 248  
QY 88 uGlyGlnMetLeuAspGluLeuAlaIleLeuTyrValLeuMetCysAlaLeuAlaMetTr 108  
Db 249 GGGCCAGCTGCTGGAGAGATCCCATCTCTGTCCTGGCGAGTGGCTATAGCATATG 308  
QY 108 pPheProArgTyrLeuPheValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLe 88  
Db 309 GATGATGCTCTGATGCACCCGATGATGCCAGAGCGCTCCCGCTACATTACGTTGTCIG 188  
QY 68 pThrLeuLeuValValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLe 88  
Db 189 GGTCTCTTCATGATCATAGGCTGTCTCCATGATATTCACATGATCCCTTCTTATCTTCGCGGCAC 248  
QY 88 uGlyGlnMetLeuAspGluLeuAlaIleLeuTyrValLeuMetCysAlaLeuAlaMetTr 108  
Db 249 GGGCCAGCTGCTGGAGAGATCCCATCTCTGTCCTGGCGAGTGGCTATAGCATATG 308  
QY 108 pPheProArgTyrLeuPheValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLe 88  
Db 309 GATGATGCTCTGATGCACCCGATGATGCCAGAGCGCTCCCGCTACATTACGTTGTCIG 188  
QY 128 lValCysValLeuSerAlaIleThrCysLeuAlaPheIleLeuProAlaIleAsnAs 148  
Db 369 GGTCTCTTCATGATCATAGGCTGTCTCCATGATATTCACATGATCCCTTCTTATCTTCGCGGCAC 248  
QY 148 nIleSerLeuMetIleLeuGlyLeuProCysThrAlaLeuValAlaGluLeuLysAr 168  
Db 429 CTACGCCCTCAACAGCATGCTCCCTGCACATCTCTACATCGTGTGCCAGAGTACAGGA 488  
QY 168 gCysAspAsnValArgValPheLeuGlyLeuPheSerGlyLeuTyrThrIleAl 188  
Db 489 GACCAGCAATAGGAGCTTCGGCACCTGATTGAGGTCTCCGTTTATGCGCTGTTGC 548

Alignment Scores: 4.89e-43 Length: 795  
Pred. No.: 549.50 Matches: 103  
Score: 59.38% Conservative: 49  
Percent Similarity: 40.23% Mismatches: 102  
Best Local Similarity: 36.34% Indels: 2  
Query Match: 9 Gaps: 1  
DB: 1

US-10-017-410-2 (1-275) x AF347024 (1-795)

QY 9 HisLeuArgAlaGly-SerSerGluValAspTyrCysGluAspAsnTyrThrIleValPr 28  
Db 9 CATCTTCGCTATCAGAGCTCCGAGGTGGACTGGTGTGAGAGCAACTCCAGTACTCGGA 68  
QY 28 oAlaIleAlaGluPheThrIleSerAsnValLeuPhePheIleLeuProProl 48  
Db 69 GCTGGTGGCGAGTCTTACACAGCTTCTCCAAATATCCCTTCTTATCTTCGCGGCAC 128  
QY 48 eCysMetCysLeuPheArgGlnTyrAlaThrCysPheAsnSerGlyIleTyrLeuIleTr 68  
Db 129 GATGATGCTCTGATGCACCCGATGATGCCAGAGCGCTCCCGCTACATTACGTTGTCIG 188  
QY 68 pThrLeuLeuValValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLe 88  
Db 189 GGTCTCTTCATGATCATAGGCTGTCTCCATGATATTCACATGATCCCTTCTTATCTTCGCGGCAC 248  
QY 88 uGlyGlnMetLeuAspGluLeuAlaIleLeuTyrValLeuMetCysAlaLeuAlaMetTr 108  
Db 249 GGGCCAGCTGCTGGAGAGATCCCATCTCTGTCCTGGCGAGTGGCTATAGCATATG 308  
QY 108 pPheProArgTyrLeuPheValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLe 88  
Db 309 GATGATGCTCTGATGCACCCGATGATGCCAGAGCGCTCCCGCTACATTACGTTGTCIG 188  
QY 128 lValCysValLeuSerAlaIleThrCysLeuAlaPheIleLeuProAlaIleAsnAs 148  
Db 369 GGTCTCTTCATGATCATAGGCTGTCTCCATGATATTCACATGATCCCTTCTTATCTTCGCGGCAC 248  
QY 148 nIleSerLeuMetIleLeuGlyLeuProCysThrAlaLeuValAlaGluLeuLysAr 168  
Db 429 CTACGCCCTCAACAGCATGCTCCCTGCACATCTCTACATCGTGTGCCAGAGTACAGGA 488  
QY 168 gCysAspAsnValArgValPheLeuGlyLeuPheSerGlyLeuTyrThrIleAl 188  
Db 489 GACCAGCAATAGGAGCTTCGGCACCTGATTGAGGTCTCCGTTTATGCGCTGTTGC 548

RESULT 11  
AF347024  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 795)  
Mao, C., Xu, R. and Obeid, L.M.

```

QY 188 aLeuPheCysTrpIleSerAspGlnAlaPheCysGluLeuLeuSerSerPheHisPhePr 208
    |||||
Db 549 TCTCACCAGCTGATCAGTACCGCTTCGTTTCAGCTTCTGGCAGAGGATTCAATTTCTT 608
QY 208 oTyLeuHisCysValTrpHisIleLeuLeuLeuAlaSerTyLeuGlyCysValCy 228
    |||||
Db 609 CTATCTGCACGATCTGGCATGTGCTCATCAGCATCACCTTCCCTTATGGCATGTCAC 668
QY 228 sPheAlaTyPheAspAlaAlaSerGluIleProGluGlnGlyProValIleArgPheTr 248
    |||||
Db 669 CATGGCCTTGGTGCATGCCAACTATGATGATGAGTGCAGGTGAACCCCTCAAGTCCGCTACTG 728
QY 248 pProSerGluIlyTrpAlaPheIleGlyValProTyTyValSerLeu 263
    |||||
Db 729 GCCTCGGGACAGTTGGCCC---GTGGGGCTGCCCTACGTGGAATC 771

RESULT 12
LOCUS CO744096 315 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 30030 from Patent WO02068579.
ACCESSION CO744096
VERSION CO744096.1 GI:42358801
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof
JOURNAL Patent: WO 02068579-A 30030 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source 1. .315
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 2.87e-43 Length: 315
Score: 547.00 Matches: 95
Percent Similarity: 99.04% Conservative: 8
Best Local Similarity: 91.35% Mismatches: 1
Query Match: 36.18% Indels: 0
DB: 6 Gaps: 0

US-10-017-410-2 (1-275) x CO744096 (1-315)
QY 172 ValArgValPheLysLeuGlyLeuPheSerGlyLeuTrpTrpThrLeuAlaLeuPheCys 191
    :|||
Db 1 ATGCGTGTGTTTAAGCTGGGCTCTTCTCGGGCTCTGTTGGAGCCCTGGCCCTGTTCTGC 60
QY 192 TrpIleSerAspGlnAlaPheCysGluLeuLeuSerSerPheHisPheProTyLeuHis 211
    |||||
Db 61 TGGATCAGTGACCGAGCTTTCTGGAGCTGTGTCTTCGAGCCCTGGCCCTGTTCTGC 120
QY 212 CysValTrpHisIleLeuLeuLeuAlaSerTyLeuGlyCysValCysPheAlaTy 231
    |||||
Db 121 TGCATCAGTGACCGAGCTTTCTGGAGCTGTGTCTTCGAGCCCTGGCCCTGTTCTGC 180
QY 232 PheAspAlaAlaSerGluIleProGluGlnGlyProValIleArgPheTrpProSerGlu 251
    |||||
Db 181 TTTGATGCTGCTCAGAGATTCTCGAGCAAGGCCCTGTCTCATCAAAATTTCTGCCCCAGGAG 240
QY 252 LysTrpAlaPheIleGlyValProTyTyValSerLeuLeuCysAlaHisLysLysSerPro 271
    |||||
Db 241 AAATGGGCCCTTCAATGTGTTGCCCTATGTGTCCCTCTGTGTGCCAACAGAAATCATCA 300
QY 272 ValIlyIleThr 275
    |||||
Db 301 GTCAAGACCACG 312

RESULT 14
LOCUS CO736906 315 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 22840 from Patent WO02068579.
ACCESSION CO736906
VERSION CO736906.1 GI:42333767
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof
JOURNAL Patent: WO 02068579-A 22839 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source 1. .315
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 4.45e-43 Length: 315
Score: 545.00 Matches: 95
Percent Similarity: 98.08% Conservative: 7
Best Local Similarity: 91.35% Mismatches: 2
Query Match: 36.04% Indels: 0
DB: 6 Gaps: 0

US-10-017-410-2 (1-275) x CO736905 (1-315)
QY 172 ValArgValPheLysLeuGlyLeuPheSerGlyLeuTrpTrpThrLeuAlaLeuPheCys 191
    :|||
Db 1 ATGCGTGTGTTTAAGCTGGGCTCTTCTCGGGCTCTGTTGGAGCCCTGGCCCTGTTCTGC 60
QY 192 TrpIleSerAspGlnAlaPheCysGluLeuLeuSerSerPheHisPheProTyLeuHis 211
    |||||
Db 61 TGGATCAGTGACCGAGCTTTCTGGAGCTGTGTCTTCGAGCCCTGGCCCTGTTCTGC 120
QY 212 CysValTrpHisIleLeuLeuLeuAlaSerTyLeuGlyCysValCysPheAlaTy 231
    |||||
Db 121 TGCATGTTGGACATCCTCATCTGCTTGTCTTACCTTCCCTTCCCTTGTATGCTTGTCTAC 180
QY 232 PheAspAlaAlaSerGluIleProGluGlnGlyProValIleArgPheTrpProSerGlu 251
    |||||
Db 181 TTTGATGCTGCTCAGAGATTCTCGAGCAAGGCCCTGTCTCATCAAAATTTCTGCCCCAGGAG 240
QY 252 LysTrpAlaPheIleGlyValProTyTyValSerLeuLeuCysAlaHisLysLysSerPro 271
    |||||
Db 241 AAATGGGCCCTTCAATGTGTTGCCCTATGTGTCCCTCTGTGTGCCAACAGAAATCATCA 300
QY 272 ValIlyIleThr 275
    |||||
Db 301 GTCAAGACCACG 312

RESULT 14
LOCUS CO736906 315 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 22840 from Patent WO02068579.
ACCESSION CO736906
VERSION CO736906.1 GI:42333767
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof
JOURNAL Patent: WO 02068579-A 22839 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source 1. .315
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 4.45e-43 Length: 315
Score: 545.00 Matches: 95
Percent Similarity: 98.08% Conservative: 7
Best Local Similarity: 91.35% Mismatches: 2
Query Match: 36.04% Indels: 0
DB: 6 Gaps: 0

US-10-017-410-2 (1-275) x CO736905 (1-315)
QY 172 ValArgValPheLysLeuGlyLeuPheSerGlyLeuTrpTrpThrLeuAlaLeuPheCys 191
    :|||
Db 1 ATGCGTGTGTTTAAGCTGGGCTCTTCTCGGGCTCTGTTGGAGCCCTGGCCCTGTTCTGC 60
QY 192 TrpIleSerAspGlnAlaPheCysGluLeuLeuSerSerPheHisPheProTyLeuHis 211
    |||||
Db 61 TGGATCAGTGACCGAGCTTTCTGGAGCTGTGTCTTCGAGCCCTGGCCCTGTTCTGC 120
QY 212 CysValTrpHisIleLeuLeuLeuAlaSerTyLeuGlyCysValCysPheAlaTy 231
    |||||
Db 121 TGCATGTTGGACATCCTCATCTGCTTGTCTTACCTTCCCTTCCCTTGTATGCTTGTCTAC 180
QY 232 PheAspAlaAlaSerGluIleProGluGlnGlyProValIleArgPheTrpProSerGlu 251
    |||||
Db 181 TTTGATGCTGCTCAGAGATTCTCGAGCAAGGCCCTGTCTCATCAAAATTTCTGCCCCAGGAG 240
QY 252 LysTrpAlaPheIleGlyValProTyTyValSerLeuLeuCysAlaHisLysLysSerPro 271
    |||||
Db 241 AAATGGGCCCTTCAATGTGTTGCCCTATGTGTCCCTCTGTGTGCCAACAGAAATCATCA 300
QY 272 ValIlyIleThr 275
    |||||
Db 301 GTCAAGACCACG 312

```



```

        |||||   :::||| |||:::|||||   :::   |||   |||
Db    640 CTGCACAGCATTTGGCACGTCCTCATAGCATCACATTTCCTTATGGTATCGTGACCATG 699
Qy    230 AlaTyrPheAspAlaAlaSerGluIleProGluGlnGlyProValIleArgPheTyrPro 249
        |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db    700 GCCCTGGTGATGCAAAGTATGAGATGCCAGATATAAACCCCTCAAAGTCCACTACTGGCCC 759
Qy    250 SerGluLysTrpAlaPheIleGlyValProTyrValSerLeu 263
        :::   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db    760 CGGACAGCTGGGTC---ATCGGGCTACCCCTATGTGGAGATC 798

```

Search completed: November 21, 2004, 08:54:39  
Job time : 3506 secs

*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 20, 2004, 21:28:52 ; Search time 433 Seconds

(without alignments)  
3333.928 Million cell updates/sec

Title: US-10-017-410-2

Perfect score: 1512  
Sequence: 1 MGAPHWDLHRAGSEVDWC.....IGVPYVSLCAHKKSPVKIT 275

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US10017410/runat\_19112004\_133619\_18988/app.query.fasta\_1.455  
-DB=N Geneseq\_23Sep04 -CPM=fastap -SUFFIX=ring -MINMATCH=0\_1 -LOOPCT=0  
-LOOPEXT=0 -UNITS=bases -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10017410 @CGN 1 1 708 @runat\_19112004\_133619\_18988 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXI=7

Database : N Geneseq\_23Sep04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID       | Description        |
|------------|--------|-------------|--------|-------------|--------------------|
| 1          | 1512   | 100.0       | 4175   | 6 ABS54148  | Ab854148 Mouse cDN |
| 2          | 1433   | 94.8        | 4202   | 8 ACC50961  | ACC50961 Human bla |
| 3          | 1433   | 94.8        | 4212   | 11 ADN38833 | Adn38833 Cancer/an |
| 4          | 1404   | 92.9        | 828    | 6 ABS54149  | Ab54149 Human cDN  |
| 5          | 1373.5 | 90.8        | 822    | 4 AAH48641  | Aah48641 Human cer |
| 6          | 1110.5 | 73.4        | 1527   | 10 ADL06640 | Adl06640 Human 3T3 |

#### ALIGNMENTS

RESULT 1  
ABS54148  
ID ABS54148 standard; cDNA; 4175 BP.  
XX  
AC ABS54148;  
XX  
DT 25-NOV-2002 (first entry)  
XX  
DE Mouse cDNA encoding a liver tumour marker protein, CRG-L1.  
XX  
KW Mouse; ss; gene; liver cancer; liver tumour; CRG-L1;  
KW hepatocellular cancer.  
XX  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers  
FT CDS 35..862  
FT /\*tag= a  
FT /product= "Liver tumour marker protein, CRG-L1"  
XX  
FN US2002115094-A1.  
XX  
PD 22-AUG-2002.  
XX  
PF 14-DEC-2001; 2001US-00017410.  
XX  
PR 14-DEC-2000; 2000US-0255674P.  
XX  
PA (FARN/) FARNHAM P J.  
PA (GRAV/) GRAVEEL C R.

|    |        |      |        |    |          |
|----|--------|------|--------|----|----------|
| 7  | 1110.5 | 73.4 | 1527   | 10 | ADL06642 |
| 8  | 1087   | 71.9 | 1623   | 10 | ADI21948 |
| 9  | 584    | 38.6 | 487    | 10 | ADI21468 |
| 10 | 568    | 37.6 | 1215   | 4  | AAK51566 |
| 11 | 568    | 37.6 | 1631   | 4  | AAK52550 |
| 12 | 549.5  | 36.3 | 792    | 4  | AAH48639 |
| 13 | 455    | 30.1 | 537    | 4  | ABL24043 |
| 14 | 320    | 21.2 | 4758   | 4  | ABL24042 |
| 15 | 264.5  | 17.5 | 4071   | 4  | ABL30432 |
| 16 | 209.5  | 13.9 | 801    | 4  | AAH48640 |
| 17 | 209.5  | 13.9 | 1063   | 6  | ABZ11993 |
| 18 | 209.5  | 13.9 | 1063   | 12 | ADM44511 |
| 19 | 208.5  | 13.8 | 2744   | 3  | AAAG6499 |
| 20 | 199    | 13.2 | 1259   | 3  | AAC40178 |
| 21 | 186    | 12.3 | 35425  | 10 | ADC87616 |
| 22 | 182    | 12.0 | 2058   | 4  | AAH15072 |
| 23 | 182    | 12.0 | 4646   | 5  | ABV27900 |
| 24 | 182    | 12.0 | 4646   | 5  | ABV28037 |
| 25 | 182    | 12.0 | 4646   | 5  | ABV22202 |
| 26 | 182    | 12.0 | 4646   | 5  | ABV22064 |
| 27 | 161    | 10.6 | 636    | 6  | ABL90053 |
| 28 | 159    | 10.5 | 639    | 5  | AAST9600 |
| 29 | 143.5  | 9.5  | 1148   | 3  | AAF13013 |
| 30 | 142    | 9.4  | 748    | 4  | AAH07955 |
| 31 | 136    | 9.0  | 354    | 10 | ABX74425 |
| 32 | 123.5  | 8.2  | 1498   | 10 | ADC87192 |
| 33 | 118.5  | 7.8  | 6755   | 2  | AAV21511 |
| 34 | 116.5  | 7.7  | 5325   | 4  | ABL23204 |
| 35 | 113    | 7.5  | 2842   | 4  | ABL23205 |
| 36 | 112    | 7.4  | 1501   | 10 | ADC87290 |
| 37 | 111    | 7.3  | 2122   | 5  | AA80960  |
| 38 | 111    | 7.3  | 2521   | 5  | AA86068  |
| 39 | 111    | 7.3  | 50335  | 9  | AA858280 |
| 40 | 111    | 7.3  | 226475 | 9  | AA858279 |
| 41 | 110.5  | 7.3  | 349981 | 10 | ADC87619 |
| 42 | 110.5  | 7.3  | 349989 | 10 | ADC86916 |
| 43 | 110    | 7.3  | 1147   | 8  | ABZ52533 |
| 44 | 109    | 7.2  | 1282   | 10 | ADC87232 |
| 45 | 108.5  | 7.2  | 5842   | 12 | ADJ12462 |

|          |           |
|----------|-----------|
| Adl06642 | Human 3T3 |
| Adi21948 | Novel hum |
| Adi21468 | Novel hum |
| Aak51566 | Human pol |
| Aak52550 | Human pol |
| Aah48639 | Human cer |
| Ab124043 | Drosophil |
| Ab124042 | Drosophil |
| Ab130432 | Drosophil |
| Aah48640 | Human cer |
| Abz11993 | Human pol |
| Adm44511 | Novel hum |
| Aag96499 | cDNA enco |
| Aac40178 | Arabidops |
| Adc87616 | Human GPC |
| Aah15072 | Human cDN |
| Abv27900 | Human pro |
| Abv28037 | Human pro |
| Abv22202 | Human pro |
| Abv22064 | Human pro |
| Ab190053 | Human pol |
| Aas79600 | DNA encod |
| Aaf13013 | Aspergill |
| Aah07955 | Human cDN |
| Abx74425 | Human cDN |
| Adc87192 | Human GPC |
| Aav21511 | Staphyloc |
| Ab123204 | Drosophil |
| Ab123205 | Drosophil |
| Adc87290 | Human GPC |
| Aa80960  | DNA encod |
| Aa86068  | DNA encod |
| Ad858280 | Human tum |
| Aa858279 | Human tum |
| Adc87619 | Human GPC |
| Adc86916 | Human GPC |
| Abz52533 | Aspergill |
| Adc87232 | Human GPC |
| Adj12462 | DNA fragm |

XX Farnham PJ, Graveel CR;  
 PI WPI; 2002-706409/76.  
 XX P-PSDB; ABG32880.  
 DR  
 XX Novel polypeptide designated as CRG-L1, useful as diagnostic marker for  
 PT liver cancer, is differentially expressed in liver tumors relative to  
 PT normal liver tissues.  
 XX  
 PS Claim 2; Page 4-7; lipp; English.  
 XX The invention relates to a polypeptide designated as CRG-L1, which is  
 CC differentially expressed in liver tumors relative to normal expression  
 CC in normal liver tissues, designated CRG-L1. Also included are the  
 CC encoding polynucleotides (in the case of the human sequence, mapping to  
 CC chromosome 9p), expression constructs, host cells, anti-CRG-L1  
 CC antibodies, identifying modulators of CRG-L1, and the use of the CRG-L1  
 CC sequence in the diagnosis of hepatocellular cancer in tumour cells from a  
 CC liver of a human or non-human animal. The CRG-L1 protein and  
 CC polynucleotide are useful as diagnostic markers for a liver cancer in  
 CC humans and non-human animals, and as a system for assessing putative  
 CC therapeutic agents. The present sequence encodes mouse CRG-L1  
 XX  
 SQ Sequence 4175 BP; 962 A; 971 C; 1012 G; 1230 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 2.98e-142 Length: 4175  
 Score: 1512.00 Matches: 275  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0

US-10-017-410-2 (1-275) x ABS54148 (1-4175)

QY 1 MetGlyAlaProHisTrpTrpAspHisLeuArgAlaGlySerSerGluValAspTrpCys 20  
 DB 35 ATGGGGCCCGGACCTGGTGGGACACCTGGGGCTGGGAGTTCGGAGGTGGATTGGTGC 94  
 QY 21 GluAspAsnTrpThrIleValProAlaIleAlaGluPheTrpAsnThrIleSerAsnVal 40  
 DB 95 GAGGACCACTACTATCGTCCCTGGCCATTCGCGAGTCTTACAAACAGATCAGCAACGTC 154  
 QY 41 LeuPhePheIleLeuProPheCysMetCysLeuPheArgGlnTrpAlaThrCysPhe 60  
 DB 155 TTGTTTTTCAATTTACTCCCATCTGCATGCTGCTGTCGCGAGTACGCAACGTCCTTC 214  
 QY 61 AsnSerGlyIleTrpLeuIleTrpThrLeuLeuValValGlyIleGlySerValTrp 80  
 DB 215 AACAGCGGCATCTACTTATATGAGCGCTCCTAGTTGTAGTGGGATTGGATCTGTCTAC 274  
 QY 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTrpVal 100  
 DB 275 TTCCATGCAACGGTGGTTCCTGGGTGAGTGTGCTGCTGCTGCAATTACACGTCGTGGCG 334  
 QY 101 LeuMetCysAlaLeuAlaMetTrpPheProArgArgTrpLeuProIlePheArgAsn 120  
 DB 335 CTGATGTGTGCTTTGGCCAGTGGTGTCCAGAGGTATTTACCAAGATCTTTCGGAAT 394  
 QY 121 AspArgGlyArgPheLeuAlaValValCysValLeuSerAlaIleThrTrpCysLeuAla 140  
 DB 395 GACAGGGGCGAGTTCAAGGCGAGTGGTGTGCTGCTGCTGCAATTACACGTCGTGGCG 454  
 QY 141 PheIleLysProAlaIleAsnAniIleSerLeuMetIleLeuGlyLeuProCysThrAla 160  
 DB 455 TTATCAACCGCCCATCAACATATTTCCCTGATGATTCTGGGACTTCATGCACTGGG 514  
 QY 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180  
 DB 515 CTGCTTGTGGAGAGTGAAGAGTGTGACATGTGCGTGTGTTTAAGCTGGGCTCTTC 574  
 QY 181 SerGlyLeuTrpTrpThrLeuAlaLeuPheCysTrpIleSerAspGlnAlaPheCysGlu 200

DB 575 TCTGGCCCTCGTGGACTCTGGCTCTCTCTCTGGATCAGCAGCAGCAGCAGCAGCAGCAG 634  
 QY 201 LeuLeuSerSerPheHisPheProTrpTrpLeuHisCysValTrpHisIleLeuIleCysLeu 220  
 DB 635 CTGCT 694  
 QY 221 AlaSerTrpLeuGlyCysValCysPheAlaTrpPheAspAlaIleSerGluIleProGlu 240  
 DB 695 GCTTCGTACCTGGGCTGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 754  
 QY 241 GlnGlyProValIleArgPheTrpProSerGluLysTrpAlaPheIleGlyValProTrp 260  
 DB 755 CAAGGTCAGTCTATCAGATTCTGGCCAGCAGAAATGGGCTTTTATTTGGTGTCCCTTAT 814  
 QY 261 ValSerLeuLeuCysAlaHisLysLysSerProValLysIleThr 275  
 DB 815 GTGTCT 859

RESULT 2  
 ACC50961  
 ID ACC50961 standard; cDNA; 4202 BP.  
 XX  
 AC ACC50961;  
 XX  
 DT 12-JUN-2003 (first entry)  
 XX  
 DE Human bladder cancer associated cDNA sequence SEQ ID NO:21.  
 XX  
 KW Human; bladder cancer; cytostatic; gene therapy; vaccine; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003003906-A2.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 03-JUL-2002; 2002WO-US021338.  
 XX  
 PR 03-JUL-2001; 2001US-0302814P.  
 PR 03-AUG-2001; 2001US-0310099P.  
 PR 08-NOV-2001; 2001US-0343705P.  
 PR 13-NOV-2001; 2001US-0350666P.  
 PR 12-APR-2002; 2002US-0372246P.  
 XX  
 PA (E0SB-) EOS BIOTECHNOLOGY INC.  
 XX  
 PI Mack DH, Aziz N;  
 XX  
 DR WPI; 2003-201532/19.  
 DR P-PSDB; ABR48156.  
 XX  
 PT Detecting a bladder cancer-associated transcript in a cell from a  
 PT patient, comprises contacting a biological sample from the patient with a  
 PT bladder cancer-associated polynucleotide or antibody.  
 XX  
 PS Claim 6; Page 235; 307pp; English.  
 XX  
 CC The present invention describes a method for detecting a bladder cancer-  
 CC associated transcript in a cell from a patient. The method comprises  
 CC contacting a biological sample from the patient with a polynucleotide  
 CC that selectively hybridizes to a sequence that is 80 % identical to a  
 CC table of sequences (see ACC50951 to ACC51059). ACC50951 to ACC51059  
 CC encode the human bladder cancer-associated proteins given in ABR48146 to  
 CC ABR48242). Bladder cancer-associated sequences from the present invention  
 CC have cytostatic activities, and can be used in antisense gene therapy and  
 CC in vaccine production. The method can be used for detecting a bladder  
 CC cancer-associated transcript in a cell from a patient. The method is  
 CC useful in diagnosing or treating bladder cancer and in screening for  
 CC compounds that modulate bladder cancer, such as hormones or antibodies.  
 CC The nucleic acid molecules from the present invention may be used in  
 CC various screening and diagnostic methods, and for gene therapy, vaccine  
 CC and/or antisense/inhibition applications

XX Sequence 4202 BP; 1045 A; 892 C; 927 G; 1338 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 2 9e-134 Length: 4202  
Score: 1433.00 Matches: 256  
Percent Similarity: 98.18% Conservative: 14  
Best Local Similarity: 93.09% Mismatches: 5  
Query Match: 94.78% Indels: 0  
DB: 8 Gaps: 0

US-10-017-410-2 (1-275) x ACC50961 (1-4202)

```
Qy 1 MetGlyAlaProHisTrpTrpAspHisLeuArgAlaGlySerGluValAspTrpCys 20
Db 77 ATGGGGCCCGCCGACCTGGTGGGACCACTGCGAGCTGTAGCTGGAGGTGGACTGGTGC 136
Qy 21 GluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrAsnThrIleSerAsnVal 40
Db 137 GAGGACAACCTACCACTCGTGGCTGCTATCGCGAGTCTTACACAGATCAGCAATGTC 196
Qy 41 LeuPhePheIleLeuProProIleCysMetCysLeuPheArgGlnTyrAlaThrCysPhe 60
Db 197 TTATTTTCATTTACGCCCACTGATGCTGTGTTGTCAGTATGCAACATGCTTC 256
Qy 61 AsnSerGlyIleTyrLeuIleTyrThrLeuLeuValValGlyIleGlySerValTyr 80
Db 257 AACAGTGGCATCTACTTAATCTGACTCTTTTGGTGTAGTGGGAATGGATCGCTCTAC 316
Qy 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTyrVal 100
Db 317 TTCATGCAACCCCTTAGTTCTTGGGTCAGATGCTTGTATGAATGTCAGTCTTGGGTT 376
Qy 101 LeuMetCysAlaLeuAlaMetTyrPheProArgArgTyrLeuProLysIlePheArgAsn 120
Db 377 CTGATGTGCTTTGGCCATCTGTTGCCAGAGGTATCTACCAAGATCTTTCGAAT 436
Qy 121 AspArgGlyArgPheLysAlaValValCysValLeuSerAlaIleThrThrCysLeuAla 140
Db 437 GACCGGGGTAGGTTCAAGGTGGTGGTTCAGTGTCTCTGCGTTACGAGTGGCTGGCA 496
Qy 141 PheIleLysProAlaIleAsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla 160
Db 497 TTTGTCAAGCCTGCATCAACACATCTCTGTATGACCCCTGGAGTTCCTTGCACGTGA 556
Qy 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180
Db 557 CTGCTCATCGCAGAGCTAAGAGGTGTGACACATGTGTGTGTTAAGTGGGCTCTTTC 616
Qy 181 SerGlyLeuTyrTrpThrLeuAlaLeuPheCysTrpIleSerAspGlnAlaPheCysGlu 200
Db 617 TCGGGCCCTGTGGTGGACCTGTCTGCTGGATCAGTGACCGAGCTTTCTGCGAG 676
Qy 201 LeuLeuSerSerPheHisPheProTyrLeuHisCysValTyrHisIleLeuIleCysLeu 220
Db 677 CTGCTGTCACTCTTCAACTCTCCCTACTGCACTGCATGTGGCAGATCCTCATCTGCTT 736
Qy 221 AlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAlaAlaSerGluIleProGlu 240
Db 737 GCTGCTACCTGGGCTGTGTATGCTTTGCCCTACTTTGATGTGCTGCTCAGAGATTCCTGAG 796
Qy 241 GlnGlyProValIleArgPheTyrProSerGluLysTrpAlaPheIleGlyValProTyr 260
Db 797 CAAGGCCCTGTGCATCAAGTCTCGGCCCAATGAGAAATGGGCTTCATTGGTGGTCCCTAT 856
Qy 261 ValSerLeuLeuCysAlaHisLysLysSerProValLysIleThr 275
Db 857 GTGTCCCTCTGTGTGGCAACAGAAATCATCAGTCAAGATCAGC 901
```

RESULT 3

ADN38833

ID ADN38833 standard; cDNA; 4212 BP.

XX

AC

XX

DT

XX

XX

DE

XX

KW

KW

KW

KW

KW

KW

KW

OS

XX

XX

PN

XX

PD

XX

XX

PF

XX

XX

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

PR

ADN38833;

17-JUN-2004 (first entry)

Cancer/angiogenesis/fibrosis-related nucleic acid, SEQ ID NO:151.

Human, differential expression; cancer; angiogenic disorder;

fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;

inflammatory disease; autoimmune disease;

retinal neovascularisation syndrome; scarring; uterine fibroid;

detection; diagnosis; prognosis; drug screening; drug targeting;

wound healing; contraception; cytostatic; cardiant; immunomodulatory;

vulnerary; gene therapy; vaccine; gene; ss.

Homo sapiens.

WO2003042661-A2.

22-MAY-2003.

13-NOV-2002; 2002WO-US036810.

13-NOV-2001; 2001US-0350666P.

21-NOV-2001; 2001US-0332464P.

29-NOV-2001; 2001US-0334393P.

03-DEC-2001; 2001US-0335394P.

14-DEC-2001; 2001US-0340376P.

08-JAN-2002; 2002US-0347211P.

10-JAN-2002; 2002US-0347349P.

08-FEB-2002; 2002US-0352500P.

13-FEB-2002; 2002US-0356714P.

20-FEB-2002; 2002US-0359077P.

29-MAR-2002; 2002US-0368809P.

12-APR-2002; 2002US-0370110P.

05-JUN-2002; 2002US-0386614P.

16-JUL-2002; 2002US-0396839P.

22-JUL-2002; 2002US-0397775P.

22-JUL-2002; 2002US-0397845P.

09-SEP-2002; 2002US-0409450P.

(EOSB-) EOS BIOTECHNOLOGY INC.

Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;

Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;

WPI: 2003-468649/44.

P-PSDB; ADN38834.

Determining the presence or absence of a pathological cell in a patient,

useful for diagnosing, prognosing or treating cancer, comprises detecting

a nucleic acid in a biological sample.

Claim 8; SEQ ID NO 151; 1385pp; English.

The invention relates to nucleic acids and proteins (ADN38833-ADN40064)

whose expression is upregulated or downregulated in specific cancers or

other diseases such as angiogenic or fibrotic disorders, and to methods

of determining the presence or absence of a pathological cell in a

patient by detecting a nucleic acid at least 80% identical to those of

the invention or by detecting a polypeptide of the invention. The

invention also relates to expression vectors and host cells comprising a

nucleic acid of the invention; antibodies which specifically bind a

polypeptide of the invention; use of such antibodies for drug targeting;

and methods of screening for modulators of activity or expression of the

polypeptides and nucleic acids. The nucleic acids, polypeptides,

antibodies and methods are useful for diagnosing, prognosing and treating

cancer and other conditions such as psoriasis, ischaemia, heart disease,

atherosclerosis, inflammatory diseases, autoimmune diseases, retinal

neovascularisation syndromes, scarring and uterine fibroids. They may

also be useful in wound healing and in contraception. The present

sequence represents a nucleic acid sequence of the invention.

XX

SQ Sequence 4212 BP; 1047 A; 894 C; 932 G; 1339 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 2,91e-134 Length: 4212  
Score: 1433.00 Matches: 256  
Percent Similarity: 98.18% Conservative: 14  
Best Local Similarity: 93.09% Mismatches: 5  
Query Match: 94.78% Indels: 0  
DB: 11 Gaps: 0

US-10-017-410-2 (1-275) x ADN38833 (1-4212)

```
QY 1 MetGlyAlaProHisTrpTrpAspHisLeuArgAlaGlySerSerGluValAspTrpCys 20
DB 87 ATGGGCGCCCGACCTGGTGGGACCGAGCTGGAGCTGGTAGCTCGGAGGTGGACTGGTGC 146
QY 21 GluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrAsnThrIleSerAsnVal 40
DB 147 GAGGACAACTACACCATCGTGCCTGCTATCGCGAGTTCCTACACACGATCAGCAATGTC 206
QY 41 LeuPhePheIleLeuProProIleCysMetCysLeuPheArgGlnTyrAlaThrCysPhe 60
DB 207 TTATTTTTCATTTTACCGCCCATCTGCATGCTGCTTTCGTCAGTATGCAACATGCTTC 266
QY 61 AsnSerGlyIleTyrLeuIleTyrThrLeuLeuValValGlyIleGlySerValTyr 80
DB 267 AACAGTGGCATCTACTTAATCTGGACTCTTTTGGTTGTAGTGGAAATGGATCCGCTAC 326
QY 81 PheHisAlaThrIleuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTrpVal 100
DB 327 TTCCATGCAACCTTAGTTTCTTGGGTGAGTGTGATGACTTCAGTCCCTTCCTGGTT 386
QY 101 LeuMetCysAlaLeuAlaMetTrpPheProArgArgTyrIleuProLysIlePheArgAsn 120
DB 387 CTGATGTGTCTTGGCCATGTGGTTCCCGAAGGTATCTACCAAGATCTTTTCGGAAT 446
QY 121 AspArgGlyArgPheLysAlaValValCysValLeuSerAlaIleThrThrCysLeuAla 140
DB 447 GACCGGGGTAGGTTCAAGGTGGTGGTCAAGTCTCTGCGTGTACGACGTGCGCGCA 506
QY 141 PheIleLysProAlaIleAsnAniIleSerLeuMetIleLeuGlyLeuProCysThrAla 160
DB 507 TTTGTCAAGCCTGCCATCAACAACATCTCTGTGATGACCTGGAGTTCCTTGCACTGCA 566
QY 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180
DB 567 CTGCTCATCGCAGAGTAAAGAGGTGTGACAACATGCGTGTGTTAAAGCTGGGCTCTTC 626
QY 181 SerGlyLeuTrpTrpThrLeuAlaLeuPheCysTrpIleSerAspGlnAlaPheCysGlu 200
DB 627 TCGGGCTCTGGTGGACCTGGCCCTGTTCTGCTGTGATCAGTGACCGAGCTTCTGCGAG 686
QY 201 LeuLeuSerSerPheHisPheProTyrLeuHisCysValTrpHisIleLeuLeuCysLeu 220
DB 687 CTGCTGTATCTCTCAACTCCCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 746
QY 221 AlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAlaAlaSerGluLeuProGlu 240
DB 747 GCTGCTACTCGGCTGTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 806
QY 241 GlnGlyProValIleArgPheTrpProSerGluLysTrpAlaPheIleGlyValProTyr 260
DB 807 CAAGGCCCTGTATCATCAAGTTCTGGCCCATGAGAAATGGCCCTTCATTTGGTGTCCCTAT 866
QY 261 ValSerLeuLeuCysAlaHisLysLysSerProValLysIleThr 275
DB 867 GTGTCCCTCTGTGTGCCCAACAGAAATCATCAGTCAAGATCACG 911
```

## RESULT 4

ABSS4149

ID ABSS4149 standard; cDNA; 828 BP.

XX

AC ABSS4149;

```
XX 25-NOV-2002 (first entry)
DT
XX Human cDNA encoding a liver tumour marker protein, CRG-L1.
DE
XX Human; ss; gene; liver cancer; liver tumour; CRG-L1;
KW hepatocellular cancer; chromosome 9p.
KW
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH 1..828
FT CDS
FT /*tag= a
FT /product= "Liver tumour marker protein, CRG-L1"
XX
PN US2002115094-A1.
XX
PD 22-AUG-2002.
XX
PF 14-DEC-2001; 2001US-00017410.
XX
PR 14-DEC-2000; 2000US-0255674P.
XX
PA (FARN/) FARNHAM P J.
XX (GRAV/) GRAVEEL C R.
XX Farnham PJ, Graveel CR;
PI
XX WPI; 2002-706409/76.
DR P-PSDB; ABG32881.
XX
PT Novel polypeptide designated as CRG-L1, useful as diagnostic marker for
PT liver cancer, is differentially expressed in liver tumors relative to
PT normal liver tissues.
XX
PS Claim 2; Page 8; 11pp; English.
XX
CC The invention relates to a polypeptide designated as CRG-L1, which is
CC differentially expressed in liver tumours relative to normal expression
CC in normal liver tissues, designated CRG-L1. Also included are the
CC encoding polynucleotides (in the case of the human sequence, mapping to
CC chromosome 9p), expression constructs, host cells, anti-CRG-L1
CC antibodies, identifying modulators of CRG-L1, and the use of the CRG-L1
CC sequence in the diagnosis of hepatocellular cancer in tumour cells from a
CC liver of a human or non-human animal. The CRG-L1 protein and
CC polynucleotide are useful as diagnostic markers for a liver cancer in
CC humans and non-human animals, and as a system for assessing putative
CC therapeutic agents. The present sequence encodes human CRG-L1
XX
SQ Sequence 828 BP; 157 A; 216 C; 211 G; 244 T; 0 U; 0 Other;
```

## Alignment Scores:

Pred. No.: 2,96e-132 Length: 828  
Score: 1404.00 Matches: 251  
Percent Similarity: 96.73% Conservative: 15  
Best Local Similarity: 91.27% Mismatches: 9  
Query Match: 92.86% Indels: 0  
DB: 6 Gaps: 0

US-10-017-410-2 (1-275) x ABSS4149 (1-828)

```
QY 1 MetGlyAlaProHisTrpTrpAspHisLeuArgAlaGlySerSerGluValAspTrpCys 20
DB 1 ATGGGCGCCCGACCTGGTGGGACCGAGCTGGAGCTGGTAGCTCGGAGGTGGACTGGTGC 60
QY 21 GluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrAsnThrIleSerAsnVal 40
DB 61 GAGGACAACTACACCATCGTGCCTGCTATCGCGAGTTCCTACACACGATCAGCAATGTC 120
QY 41 LeuPhePheIleLeuProProIleCysMetCysLeuPheArgGlnTyrAlaThrCysPhe 60
DB 121 TTATTTTTCATTTTACCGCCCATCTGCATGCTGCTTGTGATGAGTATGCAACATGCTTC 180
```

```
Qy 61 AsnSerGlyIleTyrLeuIleTyrThrLeuLeuValValValGlyIleGlySerValTyr 80
Db 181 AACAGTGCATCTACTTAATCTGACCTCTTTGGTGTAGTGGGAATGGATCGCTCTAC 240
Qy 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTyrVal 100
Db 241 TTCCATTTTACCCCTAGTTTCTTGSTCAGATGCTTGATGAATGTCAGTCTTTGGGTT 300
Qy 101 LeuMetCysAlaLeuAlaMetTyrPheProArgTyrLeuProLysIlePheArgAsn 120
Db 301 CTGATGTGTCTTGGCCATGTGTTCCCCAGAGGTATCTACCAAGATCTTTTCGGAAT 360
Qy 121 AspArgGlyArgPheLysAlaValValCysValLeuSerAlaIleThrThrCysLeuAla 140
Db 361 GACAGGGTAGTGTCAAGGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 420
Qy 141 PheIleLysProAlaIleAsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla 160
Db 421 TTTGTCAAGCTGCGCATCAACAATCTCTCTGATGACCTGGGAGTTCCTTGCACGTCA 480
Qy 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180
Db 481 CTGCTCATCGCAGAGCTAAAGAGGTGTGACAAATGCTGTGTTTAAAGCTGGGCTCTTC 540
Qy 181 SerGlyLeuTyrThrLeuAlaLeuPheCysTyrIleSerAspGlnAlaPheCysGlu 200
Db 541 TCGGGCTCTGCTGACCTGGCCCTGTCTGCTGGATCAGTACCCGAGCTTCTGCGAG 600
Qy 201 LeuLeuSerSerPheHisPheProTyrLeuHisCysValTyrPheHisIleLeuIleCysLeu 220
Db 601 CTGCTGTCTCTCTCAAGTCTCCCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Qy 221 AlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAlaAlaSerGluIleProGlu 240
Db 661 GCTGCTACCTGCTGGTGTGTATGCTTTCCTACTTGTATGCTGCTGCTGCTGCTGCTG 720
Qy 241 GlnGlyProValIleArgPheTyrProSerGluLysTyrAlaPheIleGlyValProTyr 260
Db 721 CAAGGCCCTGTCTCATCAAGTCTTGGCCCAATGAGAAATGGGCTTCATTTGGTGTCCCTAT 780
Qy 261 ValSerLeuLeuCysAlaHisLysLysSerProValLysIleThr 275
Db 781 GTGTCCCTCTGTGTGCCAACAGAAATCATCAGTCAAGATCAG 825
RESULT 5
AAH48641
ID AAH48641 standard; DNA; 822 BP.
AC AAH48641;
XX
XX
XX
XX 21-SEP-2001 (first entry)
DE Human ceramidase K3 DNA.
KW Ceramidase; human; K3; antiproliferative; anticancer; anti-eczema;
KW antipsoriasis; dermatological; ceramide; sphingosine; treatment; ds;
KW altered cell proliferation; cancer; neurodermatitis; eczema; cosmetic;
KW permeability barrier; psoriasis; ichthyosis; lamellar ichthyosis ICR2.
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
CDS 1..822
FT /*tag= a
FT /transl_except= (pos:382..387,aa:Cys)
FT /note= "No stop codon given"
XX
XX WO200155408-A1.
XX
XX 02-AUG-2001.
XX
XX 27-JAN-2001; 2001WO-EP000900.
```

```
PR 27-JAN-2000; 2000DE-01003293.
PR 09-MAR-2000; 2000DE-01011392.
XX
XX (MEMO-) MEMOREC MEDICAL MOLECULAR RES COLOGNE ST.
XX
XX Hofmann K, Conradt M;
XX
XX WPI; 2001-483256/52.
XX P-PSDB; AAB86365.
XX
XX New ceramidase containing specific structural motif, useful for
XX diagnosis, prevention and treatment of ceramidase defects, e.g.
XX ichthyosis, also in cosmetics.
XX
XX Claim 3; Page 17; 31pp; German.
XX
XX This invention describes novel human ceramidase (I) containing a specific
XX structural motif which has antiproliferative, anticancer, anti-eczema,
XX antipsoriasis and dermatological activity. (I) cleave ceramide, resulting
XX in formation of sphingosine. (II), or the nucleic acid (II) encoding it,
XX are used for diagnosis, prevention or treatment of diseases associated
XX with ceramide defects, particularly altered cell proliferation (cancer)
XX or altered ceramide layers on the skin (neurodermatitis, eczema,
XX psoriasis), also for targeted modification of the permeability barrier by
XX ceramidase or its activators, e.g. for transcutaneous delivery of
XX substances. Specifically they are used for diagnosis of ichthyosis,
XX particularly lamellar ichthyosis ICR2. (I) can also be used in cosmetics.
XX This sequence encodes human ceramidase K3
XX
XX Sequence 822 BP; 153 A; 225 C; 209 G; 235 T; 0 U; 0 Other;
```

## Alignment Scores:

```
Pred. No.: 3.54e-129 Length: 822
Score: 1373.50 Matches: 246
Percent Similarity: 94.91% Conservative: 15
Best Local Similarity: 94.45% Mismatches: 13
Query Match: 90.84% Indels: 1
DB: Gaps: 4
```

US-10-017-410-2 (1-275) x AAH48641 (1-822)

```
Qy 1 MetGlyAlaProHisTyrTyrPheHisLeuArgAlaGlySerSerGluValAspTyrCys 20
Db 1 ATGGGCGCCCGCACCTGGTGGGACGAGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGC 60
Qy 21 GluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrAsnThrIleSerAsnVal 40
Db 61 GAGGACAACTACACCATCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
Qy 41 LeuPhePheIleLeuProProIleCysMetCysLeuPheArgGlnTyrAlaThrCysPhe 60
Db 121 TTATTTTTCATTTTACCGCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
Qy 61 AsnSerGlyIleTyrLeuIleTyrThrLeuLeuValValValGlyIleGlySerValTyr 80
Db 181 AACAGTGCATCTACTTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 237
Qy 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTyrVal 100
Db 238 TTCCATGCAACCCCTAGTTTCTTGSTCAGATGCTTGATGAATGTCAGTCTTTGGGTT 297
Qy 101 LeuMetCysAlaLeuAlaMetTyrPheProArgTyrLeuProLysIlePheArgAsn 120
Db 298 CTGATGTGTCTTGGCCATGTGTTCCCCAGAGGTATCTACCAAGATCTTTTCGGAAT 357
Qy 121 AspArgGlyArgPheLysAlaValValCysValLeuSerAlaIleThrThrCysLeuAla 140
Db 358 GACAGGGTAGTGTCAAGGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 417
Qy 141 PheIleLysProAlaIleAsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla 160
Db 418 TTTGTCAAGCTGCGCATCAACAATCTCTCTGATGACCTGGGAGTTCCTTGGCTGCA 477
```



Qy 263 LeuLeuCysAlaHisLysLysSerProValLysIleThr 275  
 Db 1199 CTCCTGTGTGCAACAGAAATCATCATGATCAAGATCAG 1237  
 RESULT 7  
 ADL06642  
 ID ADL06642 standard; DNA; 1527 BP.  
 AC  
 XX ADL06642;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Human 3T3 cell conversion promoter Pp11646 DNA.  
 XX  
 XX 3T3 cell conversion; promoter; human; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 218..640  
 FT /\*tag= a  
 FT /product= "pp11646"  
 XX  
 PN CN1403477-A.  
 XX  
 PD 19-MAR-2003.  
 XX  
 XX 12-SEP-2001; 2001CN-00126725.  
 PF  
 XX 12-SEP-2001; 2001CN-00126725.  
 PR  
 XX (SHAN-) SHANGHAI XINSHIJI GENE TECHN DEV CO LTD.  
 PA  
 XX Gu J, Yang S;  
 PI  
 XX WPI; 2003-494226/47.  
 DR P-PSDB; ADL06641.  
 DR  
 XX  
 PT Human protein with function of promoting 3T3 cell conversion and its  
 FT coding sequence.  
 XX  
 PS Example 1; SEQ ID NO 3; 41pp; Chinese.  
 XX  
 CC This invention describes a novel human protein with 3T3 cell conversion  
 CC promoting function, polynucleotides encoding the polypeptide and the  
 CC recombinant process of producing the polypeptide. The present invention  
 CC also discloses the agonist resisting the polypeptide and its treatment  
 CC effect. The present invention also discloses the application of the  
 CC polynucleotides encoding the human protein with 3T3 cell conversion  
 CC promoting function.  
 CC  
 SQ Sequence 1527 BP; 376 A; 357 C; 383 G; 411 T; 0 U; 0 Other;  
 XX  
 Alignment Scores:  
 Pred. No.: 3.05e-102 Length: 1527  
 Score: 1110.50 Matches: 222  
 Percent Similarity: 66.20% Conservative: 13  
 Best Local Similarity: 62.54% Mismatches: 4  
 Query Match: 73.45% Indels: 116  
 DB: 10 Gaps: 2  
 US-10-017-410-2 (1-275) x ADL06642 (1-1527)

Qy 37 IleSerAsnValLeuPheIleLeuProPheLeuPheArgGlnTyr 56  
 Db 179 ATCAGCAATGCTTATTTTCAATTTACGCCCATCTGCATGTGTTTCTGTCAGTAT 238  
 Qy 57 AlaThrCysPheAsnSerGlyIleTyrLeuIleThrLeuValValGlyIle 76  
 Db 239 GCAACATGCTTCAACAGTGGCATCTACTTAATCTGCACCTCTTTGGTTGTAGTGGAAAT 298  
 Qy 77 GlySerValTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGlnLeuAla 96

Db 299 GGATCCGCTCTACTTCCATGCAACCCCTAGTTTCTTGGGTGAGATGCTTGATGAACCTTGC 358  
 Qy 97 IleLeuTrpValLeuMetCysAlaLeuAlaMetTrpPheProArgArgTyrLeuProLys 116  
 Db 359 GTCCTTTGGGTTCTGATGTGTGCTTTGGCCATGTGTTCGCCAGAGGATATACCAAG 418  
 Qy 117 IlePheArgAsnAspArgGlyArgPheLysAlaValValCysValLeuSerAlaIleThr 136  
 Db 419 ATCTTTCCGAATGACCGGGGTAGGTTCAAGGTGGTGTGCTGCTCTGTCGCGGTACG 478  
 Qy 137 ThrCysLeuAlaPheIleLysProAlaIleAsnAsnIleSerLeuMetIleLeuGlyLeu 156  
 Db 479 ACGTGCTGGCATTTGTCAAGCCTGCCATCAACACATCTCTCTGATGACCTGGAGTT 538  
 Qy 157 ProCysThrAlaLeuLeuValAlaGluLeuLys----- 167  
 Db 539 CTTTGCACTGCATGCTCATCGCAGAGCTAAAGAGGCATGAGAGGAACCCAGCAAGGAGA 598  
 Qy 167 ----- 167  
 Db 599 CACAGGAAAGTGGCCAGCAAGAGGTGGAGACAAGGTCTGACGATGATGACTCTCTGG 658  
 Qy 168 -----ArgCysAspAsnValArgValPheLysLeuGlyLeuPheSerGlyLeuTr 184  
 Db 659 ACCCCCGTGCAAGTGTGACACATGCGTGTGTTTAAAGCTGGCCCTCTTCTCGGGCTCTG 718  
 Qy 184 pTrpThrIleAlaLeuPheCysTrpIleSerAspGlnAlaPheCysGluLeuLeuSerSe 204  
 Db 719 GTGACCTGGCCCTGTTCTCTGATCAGTGACCGAGCTTTCGCGAGCTGCTGTCATC 778  
 Qy 204 rPheHisPheProTyrLeuHisCysValTrp----- 214  
 Db 779 CTTCAACTTCCCCTACTCTGCATGCTGTGTGGACGGTGCAGATGTTTTCAGAACCCACTG 838  
 Qy 214 ----- 214  
 Db 839 AAAGCACACTTTGTAAAGAGCTGAATTGACTCCAGGAGCAGGCTTAGCCGGAACGAAG 898  
 Qy 214 ----- 214  
 Db 899 GCCTGATTGTCATGCTTAGCAGGACTATTAAAGAACATTTAAAGACGGTCCCAAAATGC 958  
 Qy 214 ----- 214  
 Db 959 TCAAGATTATCATCTCTGCTATTCCAGGATGTGTGTAATCTTCATCCTTCTCTAAATA 1018  
 Qy 215 -----HisIleLeuIleCysLeuAlaSer 222  
 Db 1019 GAAGAGGCCCTGGGCTGCAACTACAGTCAGTCAGCAAGGCACATCCTCATCTGCTTGTGCC 1078  
 Qy 223 TyrLeuGlyCysValCysPheAlaTyrPheAspAlaIleSerGluIleProGluGlnGly 242  
 Db 1079 TACCTGGGCTGTGTATGCTTTGCTTACTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1138  
 Qy 243 ProValIleArgPheTrpProSerGluLysTrpAlaPheIleGlyValProTyrValSer 262  
 Db 1139 CTTGTTCATCAAGTTCTTGGCCCAATGAGAAATGGGGCTTTCATGTGGTGTCCCTATGTGTGCC 1198  
 Qy 263 LeuLeuCysAlaHisLysLysSerProValLysIleThr 275  
 Db 1199 CTCCTGTGTGCAACAGAAATCATCATGATCAAGATCAG 1237  
 RESULT 8  
 ADI21948  
 ID ADI21948 standard; cDNA; 1623 BP.  
 XX  
 AC ADI21948;  
 XX  
 DT 15-APR-2004 (first entry)  
 XX  
 DE Novel human protein cDNA #207.  
 XX  
 KW forensic; nutritional source; damaged tissue; diseased tissue;

KW myeloid cell disorder; lymphoid cell disorder;  
KW bone cartilage tissue growth; tendon tissue growth;  
KW ligament tissue growth; nerve tissue growth; regeneration; wound healing;  
KW tissue repair; tissue replacement; burn; incision; ulcer; cancer; human;  
KW ss; gene.

|          |   |  |      |
|----------|---|--|------|
| Qy       | 61  | AsnSerGlyIleTyrLeuIleTrpThrLeuLeuValValGlyIleGlySerValTyr    | 80   |
| Db       | 931   | -----  | 931  |
| Qy       | 81  | PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTrpVal | 100  |
| Db       | 931   | -----  | 931  |
| Qy       | 101   | LeuMetCysAlaLeuAlaMetTrpPheProArgTyrLeuProLysIlePheArgAsn    | 120  |
| Db       | 932   | -----AGGTATCTACCAAGATCTTTTCGGAAT                             | 958  |
| Qy       | 121   | AspArgGlyArgPheLysAlaValValCysValLeuSerAlaIleThrThrCysLeuAla | 140  |
| Db       | 959   | GACCAGGGTAGGTTCAGAGTGGTGGTCTGTGTCTGTCTGAGTTATGACGTGCTGGCA    | 1018 |
| Qy       | 141   | PheIleLysProAlaIleAsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla | 160  |
| Db       | 1019  | TTTGTCAAGCCTGCATCAACAACATCTCTCATGACCCCTGGGAGTTCCTTGGCTGCA    | 1078 |
| Qy       | 161   | LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe | 180  |
| Db       | 1079  | CTGCTCATCACAGAGCTAAGAGAGTGTCAACAATGCGTGTGTTTAAAGCTGGGCTCTTC  | 1138 |
| Qy       | 181   | SerGlyLeuTrpTrpThrLeuAlaLeuPheCysTrpIleSerAspGlnAlaPheCysGlu | 200  |
| Db       | 1139  | TCGGGCTCTGGTGGACCTGGCCCTGTTCTGTGGATCAGTGACCGAGCTTTCTGGCAG    | 1198 |
| Qy       | 201   | LeuLeuSerSerPheHisPheProTyrLeuHisCysValTrpHisIleLeuIleCysLeu | 220  |
| Db       | 1199  | CTGCTGTCACTCCCTCAACTCCCTCACTGCATGCTGTCGACATCTCTCATCTGCCTT    | 1258 |
| Qy       | 221   | AlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAlaAlaSerGluIleProGlu | 240  |
| Db       | 1259  | GCTGGCTACTGGGCTGTGTATGCTTTGGCTACTTTGATGCTGCCCTCAGAGATTCCTGAG | 1318 |
| Qy       | 241   | GlnGlyProValIleArgPheTrpProSerGluLysTrpAlaPheIleGlyValProTyr | 260  |
| Db       | 1319  | CAAGGCCCTGTCACTCAAGTTCTGGCCCAATGAGAAATGGGCTTCATTTGTTGCCCTAT  | 1378 |
| Qy       | 261   | ValSerLeuLeuCysAlaHisLysLysSerProValLysIleThr                | 275  |
| Db       | 1379  | GTGTCCCTCTGTGTGCCAACAGAAATCATCATCAGTCAAGATCACG               | 1423 |
| RESULT 9 |   |  |      |
| AD121468 |   |  |      |
| ID       | AD121468 standard; cDNA; 487 BP.  |  |      |
| XX       | AC AD121468;  |  |      |
| XX       | XX  |  |      |
| XX       | 15-APR-2004 (first entry)   |  |      |
| XX       | DE Novel human expressed sequence tag, EST #167.                          |  |      |
| XX       | forensic; nutritional source; damaged tissue; diseased tissue;            |  |      |
| KW       | myeloid cell disorder; lymphoid cell disorder;                            |  |      |
| KW       | bone cartilage tissue growth; tendon tissue growth;                       |  |      |
| KW       | ligament tissue growth; nerve tissue growth; regeneration; wound healing; |  |      |
| KW       | tissue repair; tissue replacement; burn; incision; ulcer; cancer; human;  |  |      |
| XX       | ss; gene; EST; expressed sequence tag.                                    |  |      |
| XX       | Homo sapiens.   |  |      |
| OS       | WO2003025148-A2.  |  |      |
| PN       | XX  |  |      |
| XX       | 27-MAR-2003.  |  |      |
| XX       | 19-SEP-2002; 2002WO-US029964.   |  |      |
| XX       | 19-SEP-2001; 2001US-0323739P.   |  |      |
| PR       | 13-SEP-2002; 2002US-00323739.   |  |      |
| XX       | (HYSE-) HYSEQ INC.  |  |      |
| PA       | XX  |  |      |



XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang D;  
 PI Haley-Vicente D;  
 XX  
 DR WPI: 2003-354603/33.  
 DR P-PSDB; ADI21688.  
 XX  
 PT New polynucleotides and secreted proteins, useful for treating myeloid or  
 PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve  
 PT tissue growth or regeneration, in wound healing, and in tissue repair and  
 PT replacement.  
 XX  
 PS Example 2; SEQ ID NO 719; 156pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide encoding a  
 CC polypeptide with biological activity. The polynucleotides and  
 CC polypeptides are useful in diagnostics, forensics, gene mapping,  
 CC identification of mutations responsible for genetic disorders and other  
 CC traits, to assess biodiversity, as nutritional sources or supplements.  
 CC The polynucleotides may also be used as molecular weight markers,  
 CC chromosome markers or map related gene positions, or as an antigen to  
 CC raise anti-DNA antibodies or elicit immune response. The polypeptides are  
 CC useful for raising antibodies, as markers for tissues in which the  
 CC corresponding polypeptide is expressed, for re-engineering damaged or  
 CC diseased tissues, for treating myeloid or lymphoid cell disorders, in  
 CC bone cartilage, tendon, ligament and/or nerve tissue growth or  
 CC regeneration, in wound healing, in tissue repair and replacement, in  
 CC healing of burns, incisions and ulcers, and in treating cancer. The  
 CC present sequence represents a novel human expressed sequence tag, EST.  
 XX  
 SQ Sequence 487 BP; 81 A; 138 C; 131 G; 134 T; 0 U; 3 Other;

## Alignment Scores:

Pred. No.: 1.12e-49 Length: 487  
 Score: 584.00 Matches: 105  
 Percent Similarity: 98.17% Conservativeness: 2  
 Best Local Similarity: 96.33% Mismatches: 2  
 Query Match: 38.62% Indels: 0  
 DB: 10 Gaps: 0

US-10-017-410-2 (1-275) x ADI21468 (1-487)

QY 1 MetGlyAlaProHisTyrTrpAspHisLeuArgAlaGlySerSerGluValAspTyrCys 20  
 DB 159 ATGGCGCGCCCGCACTGGTGGGACCGAGCTGCAGGCTGGTAGCTCGAGGTGGACTGGTGC 218  
 QY 21 GluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrAsnThrIleSerAsnVal 40  
 DB 219 GAGGACAACTACACATCGTCCCTGCTATCCCGAGTTCCTACACACAGTCAGCAATGTC 278  
 QY 41 LeuPhePheIleLeuProProIleCysMetCysLeuPheArgGlnTyrAlaThrCysPhe 60  
 DB 279 TTAATTTTCATTTTACCGCCCATCGCATGTGCTGTGTTCTGCTAGTATGCAATGCTTC 338  
 QY 61 AsnSerGlyIleTyrLeuIleTyrThrLeuValValValGlyIleGlySerValTyr 80  
 DB 339 AACAGTGGCATCTACTTAATCTGGACTCTTTTGGTGTGTAGTGGGAATGGATCCGCTCAC 398  
 QY 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTyrVal 100  
 DB 399 TTCATGCAACCCCTAGTTTCTTGGGTGAGATGCTTGAAGACTTGCAGTCCCTTGGGTT 458  
 QY 101 LeuMetCysAlaLeuAlaMetTyrPhe 109  
 DB 459 CTGATGTGTGCTTTAGCCATGTGGTTC 485

## RESULT 10

AAK51566  
 ID AAK51566 standard; cDNA; 1215 BP.

XX  
 AC AAK51566;  
 XX

DT 06-NOV-2001 (first entry)  
 XX Human polynucleotide SEQ ID NO 111.  
 DE  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation; ss.  
 OS  
 XX Homo sapiens.  
 XX  
 PN W0200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US004098.  
 XX  
 PR 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620325.  
 PR 01-SEP-2000; 2000US-00654936.  
 PR 15-SEP-2000; 2000US-00663561.  
 PR 20-OCT-2000; 2000US-00693325.  
 PR 30-NOV-2000; 2000US-00728422.  
 XX  
 FA (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;  
 XX  
 DR WPI: 2001-476283/51.  
 DR P-PSDB; AAM78433.  
 XX  
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 PT in diagnosis and gene therapy.  
 XX  
 PS Claim 1; Page 750-751; 6221pp; English.  
 XX  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication  
 XX  
 SQ Sequence 1215 BP; 213 A; 374 C; 309 G; 319 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1.54e-47 Length: 1215  
 Score: 568.00 Matches: 112  
 Percent Similarity: 53.97% Conservativeness: 51  
 Best Local Similarity: 37.09% Mismatches: 108  
 Query Match: 37.57% Indels: 31  
 DB: 4 Gaps: 3

US-10-017-410-2 (1-275) x AAK51566 (1-1215)

QY 2 GlyAlaProHisTyrTrp-----AspHisLeuArgAlaGly-Se 14  
 DB 22 GCGCAGCGCGCTGGTGGTGGCCACTGAGACAAGATGCTTCGCTATCAGAG 81  
 QY 14 rSerGluValAspTyrCysGluAspAsnTyrThrIleValProAlaIleAlaGluPheTy 34  
 DB 82 CTCGAGGTGGACTGGTGTGAGAGCACTTCAGTACTCGGAGCTGGTGGCGAGTTCTA 141

QY 34 rAsnThrIleSerAsnValLeuPhePheIleLeuProProIleCysMetCysLeuPheAr 54  
 Db 142 CAACACGTTCTCCAAATATCCCTTCCTTCATCTCGGCCCACTGATGATGCTCTGATGCA 201  
 QY 54 gGlnTyrAlaThrCysPheAsnSerGlyIleTyrLeuIleTyrThrLeuLeuValValVa 74  
 Db 202 CCCGATGCCCAGAGCGCTCCCGCTACATTTAGTTGTCTGGGTCTCTTCATGATCAT 261  
 QY 74 lGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspG1 94  
 Db 262 AGCCCTGTTCTCCATGATTTTCCACATGAGCTCAGCTTCCTGGCCGAGCTGCTGGACGA 321  
 QY 94 uLeuAlaIleLeuTyrValLeuMetCysAlaLeuAlaMetTyrPheProArgTyrLe 114  
 Db 322 GATCGCATCTCTGGCTCTGGCAGTGGCTATAGCATATGATGATGATGATGATGATTT 381  
 QY 114 uProLysIlePheArgAsnAspArgGlyArgPheLysAlaValValCysValLeuSerAl 134  
 Db 382 CCCCTCTCTCTGGGGGAACAGTCCAGTTTCATCCGCTGTCTTCATCACCACCTGT 441  
 QY 134 alleThrThrCysLeuAlaPheIleLysProAlaIleAsnAsnIleSerLeuMetIleLe 154  
 Db 442 GGTACAGACCTTCTGTCTCTCTGGCGCCCGCTCAGCTCAGCCCTCAACAGCAT 501  
 QY 154 uGlyLeuProCysThrAlaLeuValAlaGluLeuLysArgCysAspAsnValArgVa 174  
 Db 502 TGCCTGTGCATCTCTACATGTGTGTCAGGAGTACAGGAAGACCAAGCAATAGGAGCT 561  
 QY 174 lPheLysLeuGlyLeuPheSerGlyLeuTyrThrLeuAlaLeuPheCysTyrIleSe 194  
 Db 562 TCGGCACCTGATTGAGTCTCCGTGGTGTATGGCGTGTGCTCTGACCAAGCTGGATCAG 621  
 QY 194 rAspGlnAlaPheCysGluLeuLeuSerSerPheHisPheProTyrLeuHisCysValTr 214  
 Db 622 TGACCTGTCTGTTGTCAGCTTCTGGCAGAGGATTCATTTCTTCTATCTGCACAGCATCTG 681  
 QY 214 pHisIleLeuIleCysLeuAlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAl 234  
 Db 682 GCATGTCTCATCAGATCATCTCCCTTATGGATGATGATGATGATGATGATGATGATG 741  
 QY 234 aAlaSerGluIleProGluGlnGlyProValIleArgPheTyrProSerGluLysTyrPal 254  
 Db 742 CAACTATGATGATGCCAGGTGAACCCCTCAAGTCCGCTACTGCGCTCGGACAGTTGGCC 801  
 QY 254 aPheIleGlyValProTyrValSerLeu----- 263  
 Db 802 C---GTGGGGCTGCCCTACGTGGAAATCCGGGGTGATGACAAGGACTGCTGACACCTGCC 858  
 QY 264 -----LeuCysAlaHisLysLysSerProVa 272  
 Db 859 AGCCTCTTGACTATCCACACCCCACTTGCCTGTGTCTTGTGAGAGATAGCCCGGT 918  
 QY 272 lLys 273  
 Db 919 CAGG 922  
 RESULT 11  
 AAK52550  
 ID AAK52550 standard; cDNA; 1631 BP.  
 XX  
 AC  
 XX  
 XX  
 DT  
 DT  
 XX  
 DE  
 XX  
 KW  
 KW  
 KW  
 KW  
 XX  
 XX  
 OS

XX WO200157190-A2.  
 XX 09-AUG-2001.  
 XX 05-FEB-2001; 2001WO-US004098.  
 XX 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620325.  
 PR 01-SEP-2000; 2000US-00654936.  
 PR 15-SEP-2000; 2000US-00663561.  
 PR 20-OCT-2000; 2000US-00693325.  
 PR 30-NOV-2000; 2000US-00728422.  
 XX (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue Au, Yang Y, Wejhrman T, Goodrich R;  
 XX WPI; 2001-476283/51.  
 DR P-PSDB; AAM79417.  
 XX  
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 PT in diagnosis and gene therapy.  
 XX  
 PS Claim 1; Page 4462; 6221pp; English.  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication  
 XX  
 SQ Sequence 1631 BP; 247 A; 399 C; 344 G; 350 T; 0 U; 291 Other;  
 Alignment Scores:  
 Pred. No.: 2,26e-47 Length: 1631  
 Score: 568.00 Matches: 112  
 Percent Similarity: 53.97% Conservative: 51  
 Best Local Similarity: 37.09% Mismatches: 108  
 Query Match: 37.57% Indels: 31  
 DB: 4 Gaps: 3  
 US-10-017-410-2 (1-275) x AAK52550 (1-1631)  
 QY 2 GlyAlaProHisTyrTrp-----AspHisLeuArgAlaGly-Se 14  
 Db 22 GCGCGAGCGCGTGTGGTGGCCACTGAGACAAGATGCGCTAGCATCTCGCTATCAGAG 81  
 QY 14 rSerGluValAspTyrCysGluAspAsnTyrThrIleValProAlaIleAlaGluPheTy 34  
 Db 82 CTCCGAGGTGGACTGTGTGAGAGCAACTTCAGTACTCGGAGCTGGTGGCCGAGTCTCA 141  
 QY 34 rAsnThrIleSerAsnValLeuPhePheIleLeuProProIleCysMetCysLeuPheAr 54  
 Db 142 CAACACGTTCTCCAAATATCCCTTCCTTCATCTCGGCCCACTGATGATGATGATGAT 201  
 QY 54 gGlnTyrAlaThrCysPheAsnSerGlyIleTyrLeuIleTyrThrLeuLeuValValVa 74  
 Db 202 CCCGATGCCCAGAGCGCTCCCGCTACATTTAGTTGTCTGGGTCTCTTCATGATCAT 261  
 QY 74 lGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspG1 94



QY 188 aLeuPheCysTyrIleSerAspGlnAlaPheCysGluLeuLeuSerSerPheHisPhePr 208  
 Db 549 TCGACACGCTGGATCAGTACCGCTCGCTTTCGAGCTTCGGCAGAGGATTCATTTCTT 608  
 QY 208 oTyrLeuHisCysValTyrHisIleLeuIleCysLeuAlaSerTyrLeuGlyCysValCy 228  
 Db 609 CTAATCGACAGCATCTGGCATGTGCTCATCAGCATCACCTTCCTTATGGCATGGTAC 668  
 QY 228 sPheAlaTyrPheAspAlaSerGluIleProGluGlnGlyProValIleArgPheTr 248  
 Db 669 CATGGCCTTGGTGGATGCCAACTATGATGATGCCAGGTGAAACCCCTCAAGTCCGCTACTG 728  
 QY 248 pProSerGluLysTyrAlaPheIleGlyValProTyrValSerLeu 263  
 Db 729 GCCTCGGACAGTTGGCCC---GTGGGGCTGGCCCTACGTGGAATC 771

RESULT 13  
 ABL24043  
 ID ABL24043 standard; DNA; 537 BP.  
 AC ABL24043;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 23602.  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.  
 XX  
 PS Claim 1; SEQ ID NO 23602; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB1630511), expressed DNA  
 CC sequences (AB161840-AB16175) and the encoded proteins (AB57737-  
 CC AB572072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 537 BP; 102 A; 146 C; 148 G; 141 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 1.39e-36 Length: 537  
 Score: 455.00 Matches: 82  
 Percent Similarity: 69.14% Conservative: 30  
 Best Local Similarity: 50.62% Mismatches: 50  
 Query Match: 30.09% Indels: 0  
 DB: 4 Gaps: 0

US-10-017-410-2 (1-275) x ABL24043 (1-537)  
 QY 7 TrpAspHisLeuArgAlaGlySerSerGluValAspTyrCysGluAspAsnTyrThrIle 26  
 Db 49 TGGGAGCACCCTAGAGCCCGAGAGCTGCCGCTGACCTGGTGGAGGCACTACTTGTATT 108  
 QY 27 ValProAlaIleAlaGluPheTyrAsnThrIleSerAsnValLeuPhePheIleLeuPro 46  
 Db 109 TCGTCCAAATCGCCGAGTTCTGTGAACACGTTTAGCAACTTCTGTTCATCTACTGCG 168  
 QY 47 ProIleCysMetCysLeuPheArgGlnTyrAlaThrCysPheAsnSerGlyIleTyrLeu 66  
 Db 169 CCGTCTCTATATATGTCTTCAAGGAGTACGGACGCTTTGTGACGCCCGGAATCCAGTC 228  
 QY 67 IleTyrThrLeuLeuValValGlyIleGlySerValTyrPheHisAlaThrLeuSer 86  
 Db 229 ATCTGGGTGCTGCTCATCGTGGTGGCTGAGTTCGATGTACTTCCATGCCACTTTGAGT 288  
 QY 87 PheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTyrValLeuMetCysAlaLeuAla 106  
 Db 289 CTGATTGGCCAGCTGCTGGACGAACCTGGCCATCTCTGGGTCTTTCATGGCGGCTTTTCG 348  
 QY 107 MetTyrPheProArgTyrLeuProLysIlePheArgAsnAspArgGlyArgPheLys 126  
 Db 349 CTCTTCTATCCGAAGCGATCTATCCAGTTCGTGAAAACCATCGCAAAACCTTCAGT 408  
 QY 127 AlaValCysValLeuSerAlaIleThrThrCysLeuAlaPheIleLysProAlaIle 146  
 Db 409 TGGCTCATGCTCTTGTGGCGGATTGCTGGAGCGGCTTGTGCGGTGGAGGCCATTGTT 468  
 QY 147 AsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAlaLeuLeuValAlaGluLeu 166  
 Db 469 AACGCTTTGTCTCATGTTTCATGAGTGTGCCGACCATGATGTAATGTCTTACAGAGCTG 528  
 QY 167 LysArg 168  
 Db 529 CAGAGG 534

RESULT 14  
 ABL24042  
 ID ABL24042 standard; DNA; 4758 BP.  
 AC ABL24042;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 23599.  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.  
 XX

PS Claim 1; SEQ ID NO 23599; 2lpp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 4758 BP; 1304 A; 981 C; 980 G; 1493 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 1.07e-21 Length: 4758  
Score: 320.00 Matches: 64  
Percent Similarity: 60.00% Conservativity: 29  
Best Local Similarity: 41.29% Mismatches: 38  
Query Match: 21.16% Indels: 24  
DB: 4 Gaps: 2  
US-10-017-410-2 (1-275) x ABL24042 (1-4758)  
Qy 38 SerAsnValLeuPhePheLeuProProLeuCysMetCysLeuPheArgGlnTyrAla 57  
Db 3291 AGCAACTTCCTGTTTCATCTACTGCGCGCTCTTATATGCTCTTCAAGGAGTACGGA 3350  
Qy 58 ThrCysPheAsnSerGlyLeuTyrLeuLeuTrpThrLeuValValGlyLeuGly 77  
Db 3351 CGCTTTGTGAGCGCGGAAATCCAGTCATCTGGTGTCTGCTCATCGTGGTGGCTGAGT 3410  
Qy 78 SerValTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIle 97  
Db 3411 TCGATGACTTCATGCGCCATCTTGAGTCTGATTGGCCAGCTGCTGGACGAATCGCCATA 3470  
Qy 98 LeuTrpValLeuMetCysAlaLeuAlaMetTrpPheProArgArgTyrLeuProLysIle 117  
Db 3471 CTCTGGGTCTTCATGGCGGCTTTTCGCTCTCTTATCCGAAGCATACTATCCCAAGTTC 3530  
Qy 118 PheArgAsnAspArg----- 122  
Db 3531 GTGAATAACGATCGGTAAAGCACTGTGATCCCTGACGCTGGCACTATCTAAATCCCGCG 3590  
Qy 123 -----GlyArgPheLysAlaValValCysValLeuSerAla 134  
Db 3591 ACTTTGGTTTCCTTTGTTCCTCCCGACCAAACTTTCAGTTCGCTCATGCTCTTGTGGCG 3650  
Qy 135 Ile---ThrThrCysLeuAlaPheIleLysProAlaIleAsnAsnIleSerLeuMetIle 153  
Db 3651 ATTGCTCGAGCGGCTTTGTGCTGGTGAAGCCCATTTGTTACGGCTTTGTTCTCATGTTTC 3710  
Qy 154 LeuGlyLeuProCysThrAlaLeuLeuValAlaGluLeuLysArg 168  
Db 3711 ATGAGTGTGGCGACCATGTTAATGCTCTACACAGAGCTGCGAGG 3755  
RESULT 15  
ABL30432  
ID ABL30432 standard; DNA; 4071 BP.  
XX  
AC ABL30432;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 42769.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
XX  
OS Drosophila melanogaster.  
XX  
FN WO200171042-A2.

XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Claim 1; SEQ ID NO 42769; 2lpp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 4071 BP; 1027 A; 938 C; 941 G; 1165 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 3.55e-16 Length: 4071  
Score: 264.50 Matches: 44  
Percent Similarity: 70.83% Conservativity: 24  
Best Local Similarity: 45.83% Mismatches: 27  
Query Match: 17.49% Indels: 1  
DB: 4 Gaps: 1  
US-10-017-410-2 (1-275) x ABL30432 (1-4071)  
Qy 168 ArgCysAspAsnValArgValPheLysLeuGlyLeuPheSerGlyLeuTrpTrpThrLeu 187  
Db 3214 AGAGTTAGTACCAAGAGGTCTACCGCTGGCGATCCGATCGACCGCTCTGGGCTGT 3273  
Qy 188 AlaLeuPheCysTrpIleSerAspGlnAlaPheCysGluLeuLeuSerSerPheHisPhe 207  
Db 3274 CGGTGTTCTGCTGGATCAATGACCGGATCTTTCGAGGCGCTGCTCTCGATCAATTTT 3333  
Qy 208 ProTyrLeuHisCysValTrpHisIleLeuIleCysLeuAlaSerTyrLeuGlyCysVal 227  
Db 3334 CCATACCTGCGAGGCTTCTGGCACATATTTCATTTTATAGCGGCTACACGGTGTGGTG 3393  
Qy 228 CysPheAlaTyrPheAspAlaAlaSerGluIleProGluGlnGlyProValIleArgPhe 247  
Db 3394 CTTTTCGCTACTTCTCTACGTGGATCGAGCTGCGCCGACGACGCGCTCTCGAAGTAC 3453  
Qy 248 TrpProSerGluLysTrpAlaPheIleGlyValProTyrValSerLeu 263  
Db 3454 TGGCAAGAACGAGTTCGAGTTC---GGATACCCCTTCATTTCGATC 3498  
Search completed: November 21, 2004, 07:51:29  
Job time : 445 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 21, 2004, 07:04:58 ; Search time 98 Seconds  
(without alignments)  
1994.561 Million cell updates/sec

Title: US-10-017-410-2

Perfect score: 1512  
Sequence: 1 MGAHPWHDLRAGSEVDWC.....IGVPYVSLCAHKKSPVKIT 275

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_p/US10017410/runat 19112004.133621.19022/app query.fasta.1.455  
-DB=Issued Patents NA -OFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCI=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human4.0.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10017410 @CGN 1.1.105 @runat 19112004.133621.19022 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/1/ina/5A COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID               | Description        |
|------------|-------|-------------|--------|---------------------|--------------------|
| 1          | 209.5 | 13.9        | 1063   | US-09-799-451-875   | Sequence 875, App  |
| c 2        | 204   | 13.5        | 358    | US-09-621-976-11241 | Sequence 11241, A  |
| 3          | 173   | 11.4        | 915    | US-09-248-796A-1441 | Sequence 1441, Ap  |
| c 4        | 118.5 | 7.8         | 6755   | US-08-931-999-4     | Sequence 4, Appli  |
| 5          | 105.5 | 7.0         | 1257   | US-09-252-931A-892  | Sequence 892, App  |
| c 6        | 105.5 | 7.0         | 2859   | US-03-232-931A-944  | Sequence 944, App  |
| 7          | 102   | 6.7         | 1575   | US-09-252-991A-1139 | Sequence 1139, Ap  |
| 8          | 101.5 | 6.7         | 107820 | US-03-792-616-1     | Sequence 1, Appli  |
| 9          | 101   | 6.7         | 1794   | US-09-328-352-577   | Sequence 577, App  |
| c 10       | 98.5  | 6.5         | 3304   | US-09-710-279-4182  | Sequence 4182, Ap  |
| 11         | 97.5  | 6.4         | 960    | US-09-248-796A-2470 | Sequence 2470, Ap  |
| 12         | 97.5  | 6.4         | 92407  | US-09-596-002-36    | Sequence 36, Appli |

|      |      |     |        |   |                      |                   |
|------|------|-----|--------|---|----------------------|-------------------|
| 13   | 97   | 6.4 | 2304   | 4 | US-09-540-236-426    | Sequence 426, App |
| c 14 | 97   | 6.4 | 8155   | 4 | US-08-956-171E-63    | Sequence 63, Appl |
| c 15 | 97   | 6.4 | 8155   | 4 | US-08-781-986A-63    | Sequence 63, Appl |
| 16   | 96.5 | 6.4 | 1297   | 3 | US-09-083-521-4      | Sequence 4, Appli |
| 17   | 96.5 | 6.4 | 1863   | 1 | US-09-198-284-4      | Sequence 4, Appli |
| c 18 | 96.5 | 6.4 | 1863   | 2 | US-08-987-122-4      | Sequence 4, Appli |
| c 19 | 96.5 | 6.4 | 3010   | 4 | US-08-961-527-25     | Sequence 25, Appl |
| 20   | 96   | 6.3 | 3267   | 3 | US-09-453-702B-170   | Sequence 170, App |
| 21   | 95.5 | 6.3 | 5391   | 4 | US-08-811-519-2      | Sequence 2, Appli |
| 22   | 95.5 | 6.3 | 5693   | 4 | US-09-262-537-19     | Sequence 19, Appl |
| 23   | 95   | 6.3 | 1350   | 4 | US-09-328-352-3355   | Sequence 3355, Ap |
| 24   | 94.5 | 6.2 | 3603   | 4 | US-09-221-013A-5     | Sequence 5, Appli |
| c 25 | 94.5 | 6.2 | 3673   | 4 | US-09-221-013A-11    | Sequence 11, Appl |
| 26   | 93   | 6.2 | 774    | 4 | US-09-252-991A-3254  | Sequence 3254, Ap |
| 27   | 93   | 6.2 | 5610   | 4 | US-09-262-537-57     | Sequence 57, Appl |
| 28   | 93   | 6.2 | 246240 | 2 | US-08-724-394A-20    | Sequence 20, Appl |
| 29   | 93   | 6.2 | 246240 | 2 | US-08-724-394A-21    | Sequence 21, Appl |
| 30   | 93   | 6.2 | 246240 | 2 | US-08-724-394A-22    | Sequence 22, Appl |
| 31   | 92.5 | 6.1 | 1329   | 4 | US-09-602-787A-507   | Sequence 507, App |
| 32   | 92.5 | 6.1 | 1346   | 4 | US-09-761-962A-12    | Sequence 12, Appl |
| c 33 | 92.5 | 6.1 | 1806   | 4 | US-09-248-796A-10630 | Sequence 10630, A |
| 34   | 92   | 6.1 | 648    | 4 | US-09-583-110-774    | Sequence 774, App |
| 35   | 92   | 6.1 | 1571   | 4 | US-09-221-017B-625   | Sequence 625, App |
| c 36 | 92   | 6.1 | 1763   | 4 | US-09-270-767-5952   | Sequence 5952, Ap |
| c 37 | 92   | 6.1 | 1763   | 4 | US-09-270-767-21234  | Sequence 21234, A |
| c 38 | 92   | 6.1 | 51259  | 3 | US-08-781-891-209    | Sequence 209, App |
| c 39 | 92   | 6.1 | 51259  | 4 | US-09-618-166-209    | Sequence 209, App |
| 40   | 91.5 | 6.1 | 1452   | 4 | US-09-248-796A-6712  | Sequence 6712, Ap |
| 41   | 91.5 | 6.1 | 1733   | 4 | US-09-560-761-17     | Sequence 17, Appl |
| 42   | 91.5 | 6.1 | 2612   | 4 | US-09-270-767-13970  | Sequence 13970, A |
| c 43 | 91.5 | 6.1 | 3897   | 4 | US-09-540-236-414    | Sequence 414, App |
| c 44 | 91.5 | 6.1 | 5099   | 4 | US-09-887-052-1      | Sequence 1, Appli |
| c 45 | 91.5 | 6.1 | 5099   | 4 | US-09-887-052-3      | Sequence 3, Appli |

#### ALIGNMENTS

RESULT 1

US-09-799-451-875  
; Sequence 875, Application US/09799451

; Patent No. 6783969

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Zhou, Ping

; APPLICANT: Goodrich, Ryle

; APPLICANT: Asundi, Vinod

; APPLICANT: Ren, Feiyan

; APPLICANT: Zhang, Jie

; APPLICANT: Xue, Aidong J.

; APPLICANT: Zhao, Qing A.

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Ma, Yunqing

; APPLICANT: Yamazaki, Victoria

; APPLICANT: Chen, Rui-hong

; APPLICANT: Wang, Zhiwei

; APPLICANT: Wang, Dunrui

; APPLICANT: Yang, Yonghong

; APPLICANT: Wehrman, Tom

; APPLICANT: Ghosh, Reena

; APPLICANT: Drmanac, Radjoje T.

; TITLE OF INVENTION: No. 6783969el Nucleic Acids and

; TITLE OF INVENTION: Polypeptides

; FILE REFERENCE: 803

; CURRENT APPLICATION NUMBER: US/09/799,451

; CURRENT FILING DATE: 2001-03-05

; NUMBER OF SEQ ID NOS: 948

; SOFTWARE: pt\_FL\_Genes Version 2.0

; SEQ ID NO 875

; LENGTH: 1063

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS











```

US-10-017-410-2 (1-275) x US-09-792-616-1 (1-107820)
QY 19 TrpCysGluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrAsnThrIleSer 38
Db 38295 TTCTGTCCTTAAATGCTACTGTATG-----TTGCTGGAATTTGGCATCATGTACT 38348
QY 39 AsnValLeuPhePheIleLeuProPheIleCysMetCysLeuPheArgGlnTyrAlaThr 58
Db 38349 ATAATGTCGCTGTTTCTTACTGTTACTATGCAATGCAATGCTATTCGTTACTATATAATTC 38408
QY 59 CysPheAsnSerGlyIle-----TyrLeuIleTyrPheIleValVal 73
Db 38409 AGTTATTTTGGTGAATGTCATCATGTTACTATAATGCTGCTGCTGTTTACTGTTA 38468
QY 74 ValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLeu-----88
Db 38469 -----CATTGTTATGTTGCTACTCTGTTACTCTGTTACTGTTACAGT 38504
QY 89 -----GlyClnMetLeuAsp-----GluLeuAlaIleLeuTyrPheValLeuMetCys 103
Db 38505 CGTGTGCTGGGATGTTGCTGTCATGTTATGATGTTGCTGTTGTTACTATTGCAATGTC 38564
QY 104 AlaLeuAlaMetTrpPheProArgTyrLeuProLysIlePheArgAsnAspArgGly 123
Db 38565 TAGTGTGCTGTTGTTGTTCTATCATGTTACTA-AAA-----38599
QY 124 ArgPheLysAlaValValCysValLeuSerAlaIleThrThr-----CysLeu 139
Db 38600 -----TGTTGCTATTACGTTACTATTATGCTATGTTGTTGTTGTTATT 38644
QY 140 AlaPheIleLysProAla-----IleAsnAsnIleSerLeuMetIleLeuGlyLeuProCys 158
Db 38645 GCTGGAATGTCGCATCATGTTACTATAATGCTCTATTATGTTACTA---TTACACTGC 38701
QY 159 ---ThrAlaLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeu 177
Db 38702 TAGTGTGCTATTCGTGA-----CTGCAATATGTTGTTATGTTGCTGTT-----38746
QY 178 GlyLeuPheSerGlyLeuTyrPheThrLeuAlaLeuPheCysTrpIleSerAspGlnAla 197
Db 38747 -----ACATTACTGTTTACATTGCTGTTGCTATTCTCATG 38782
QY 198 PheCysGluLeuLeuSerSerPheHisPheProTyrLeuHisCysValTrpHisIleLeu 217
Db 38783 TTCTGGAATATTGCCAATGTTGCTGTTGCCATTA-----TGTTACTATCATGTTGCA 38836
QY 218 IleCysLeuAlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAlaAaSerGlu 237
Db 38837 GCTATGTTG-----CTGTTGCAATGC---TGTTCTTGTGTTCCCTGCACCTCCATCG 38887
QY 238 Ile-ProGlu-----GlnGlyProValIleArgPheTrpProSerGluLysTr 253
Db 38888 GTGACCTGTTCTCCTCAAGCTCAGAACGAAAGCAAGATTGCGCTGGCTCCAAA--38945
QY 253 PhePheIleGlyValProTyrValSerLeuLeuCysAlaHis 267
Db 38946 -----CCTATGCTCCTAACCACTGCTGCCAC 38972

RESULT 9
US-09-328-352-577
; Sequence 577, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 577
; LENGTH: 1794

; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-577
Alignment Scores:
Pred. No.: 0.107 Length: 1794
Score: 101.00 Matches: 56
Percent Similarity: 37.97% Conservatives: 34
Best Local Similarity: 23.63% Mismatches: 67
Query Match: 6.68% Indels: 80
DB: 4 Gaps: 10

US-10-017-410-2 (1-275) x US-09-328-352-577 (1-1794)
QY 25 ThrIleValProAlaIleAlaGluPheTyrAsnThrIleSerAsnValLeuPhePheIle 44
Db 97 AGTATTTTGTACTTAATGCTGCTAATGCTGTTGATGCTTCCAGATAGCATTCACATCA 156
QY 45 LeuProIleCysMetCysLeuPheArgGlnTyrAlaThrCysPheAsnSer-----62
Db 157 ATTCACACCTGTTATATTACTATG-----ACAGCTGGGTGAGTATCAAA 201
QY 63 -----GlyIleTyrLeuIleTyrThrLeu 70
Db 202 AAGACTTTAAGTGAGGAGCTAAGGCAAACTTTACATGGCATCTCGCGCTTAATGTCGGT 261
QY 71 LeuValValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLeuGlyGln 90
Db 262 TTGAGTTTGGCTGTTAATGCGCAAGTTTATCAACCGGTGCAGACAGTATTATTTA---315
QY 91 MetLeuAspGluLeuAlaIleLeuTyrValLeuMetCysAlaLeuAlaMetTrpPhePro 110
Db 316 -----CTTTCTTGATTTGGACTTTGCTT-----339
QY 111 ArgArgTyrLeuProLysIlePheArgAsnAspArgGlyArgPheLysAlaValCys 130
Db 340 -----TTATGCGGTGCTATATCGCAAAATATTGGAATCTTT---GCTTTAATTTGT 390
QY 131 ValLeuSerAlaIleThrThrCysLeuAlaPheIleLysPro-----144
Db 391 ATCACAAAGCAACTTACACTTTTTTTTATTTTATAGACAGACTTTTTTGGCAGAGAGTTT 450
QY 145 -----AlaIleAsnAsnIleSerLeuMet-----IleLeuGly 155
Db 451 CCTATCTTTATCTTTTCTCTTAAATCTATTAACTGTTGCGAGTTTGGATCTGTCAA 510
QY 156 LeuProCysThrAlaLeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPhe 175
Db 511 AAGAAATATACGGCTTTACGTTT-----ATTTT 540
QY 176 LysLeuGlyLeuPheSerGlyLeuTyrPheThrLeuAlaLeuPheCysTrpIleSerAsp 195
Db 541 -----ATTGCTGTTGTTGCTGTTATTTCAATA 567
QY 196 GlnAlaPheCysGluLeuLeuSerSerPheHisPheProTyrLeuHisCysValTrpHis 215
Db 568 ACAGGAATGATTCAATTCATCAAGTGAATTTACCTTATCTG-----612
QY 216 IleLeuIleCysLeuAlaSerTyrLeuGlyCysValCysPheAlaTyrPhe 232
Db 613 -----ATCTCGGCTTTTTTCTAGGAATCATCGCTTTTATTATTATTT 654

RESULT 10
US-09-710-279-4182/c
; Sequence 4182, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P034800S
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258

```



|     |    |  |     |
|-----|----|--|-----|
| 148 | Qy | nlleSerLeu-Metlle-----LeuGlyLeuProCysThrA                    | 160 |
|     |    | :  |     |
| 444 | Db | CATTGGCTTGGGTGATAACACGCTTGTGCTGATTATTTTATTATTAACAGTTTGTTCGA  | 503 |
|     |    | :  |     |
| 160 | Qy | laleuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuP | 180 |
|     |    | :  |     |
| 504 | Db | GGTGTGTTGTATGCCCCCTTCAAAATGTTTTCGTGTATGTGATATCGGGTGATCCCTAT  | 563 |
|     |    | :  |     |
| 180 | Qy | he-SerGlyLeuTriPThrLeuAlaLeuPheCysTriPleSerAspGlnAla---Phe   | 198 |
|     |    | :  |     |
| 564 | Db | TCCAATGACTCTGG-----TGTTTCATATTGATAGTTGCCAAAGAG               | 605 |
|     |    | :  |     |
| 199 | Qy | CysGluLeuLeuSerSerPheHisPheProTryLeuHisCys---ValTrpHisileLeu | 217 |
|     |    | :  |     |
| 606 | Db | TGTCGCTTTCCT-----CTTAGGTGTACCACCTGG-----                     | 635 |
|     |    | :  |     |
| 218 | Qy | IleCysLeuAlaSerTryLeuGlyCys---ValCys-PheAlaTryPheAspAlaAlase | 236 |
|     |    | :  |     |
| 636 | Db | -----TCCTTGGTGTGTGTGCAGATTGTCATTTATTTATTTAAGTTTACGA          | 683 |
|     |    | :  |     |
| 236 | Qy | rgluileProGluGlnGlyProValIleArgPheTrpProSerGluLysTrpAlaPheIl | 256 |
|     |    | :  |     |
| 684 | Db | TAGATTG-----ATTAGATT-----ATCAGTCCTTGGGCTTTAAT                | 719 |
|     |    | :  |     |
| 256 | Qy | eGlyValProTryValSerLeuLeu                                    | 264 |
|     |    | :  |     |
| 720 | Db | TGGATTGTTGTACACTATTATTGTT                                    | 744 |
|     |    | :  |     |

```

RESULT 12
US-09-596-002-36
; Sequence 36, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Iagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Beig, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 36
; LENGTH: 92407
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte template ID No. 6632636 36
; PUBLICATION INFORMATION:
; US-09-596-002-36

```

| Alignment Scores:      |        |
|------------------------|--------|
| Pred. No.:             | 93.1   |
| Score:                 | 97.50  |
| Percent Similarity:    | 39.72% |
| Best Local Similarity: | 23.83% |
| Query Match:           | 6.45%  |
| DB:                    | 4      |
| Length:                | 92407  |
| Matches:               | 51     |
| Conservative:          | 34     |
| Mismatches:            | 79     |
| Indels:                | 50     |
| Gaps:                  | 11     |

US-10-017-410-2 (1-275) X US-09-596-002-36 (1-92407)

|    |       |   |       |
|----|-------|---|-------|
| Qy | 74    | ValGlyIleClySerValTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAsp  | 93    |
|    |       |   |       |
| Db | 89280 | GTATCAATTGGCAGG---TATTTC-----GCATTTCGTGCATTCAGGAC             | 89324 |
|    |       |   |       |
| Qy | 94    | GluLeuAlaIle-LeuTrpValLeuMetCysAlaLeuAlaMetTirPheProArgGArgTy | 113   |
|    |       |   |       |
| Db | 89325 | GCATGTGCTATTTTATGCGGTGATTTTATCAGCATCTTGTAACCTTTATGACGGATAAAT  | 89384 |
|    |       |   |       |
| Oy | 113   | rLeuProLysIlePheArgAsn---AspArgGlyVArqPheLysAlaValValCysValle | 132   |

|    |       |      |     |       |      |     |       |       |     |       |     |     |       |     |     |       |     |     |       |     |     |       |     |       |       |       |       |
|----|-------|------|-----|-------|------|-----|-------|-------|-----|-------|-----|-----|-------|-----|-----|-------|-----|-----|-------|-----|-----|-------|-----|-------|-------|-------|-------|
| Db | 89385 | 132  | Db  | 89445 | 150  | Qy  | 89505 | 170   | Qy  | 89538 | 190 | Db  | 89589 | 210 | Qy  | 89610 | 230 | Db  | 89661 | 244 | Qy  | 89721 | 257 | Qy    | 89781 | Db    | 89820 |
|    | :     | :    | :   | :     | :    | :   | :     | :     | :   | :     | :   | :   | :     | :   | :   | :     | :   | :   | :     | :   | :   | :     | :   | :     | :     | :     | :     |
|    |       |      |     |       |      |     |       |       |     |       |     |     |       |     |     |       |     |     |       |     |     |       |     |       |       |       |       |
|    | TG    | CACA | ACG | CA    | TTAT | TG  | CA    | CA    | CTT | GC    | CA  | ACT | TG    | CG  | CA  | AA    | TT  | CG  | TA    | TG  | AT  | GC    | TA  | GT    | GT    | TT    | 89444 |
|    | user  | Ala  | lle | Thr   | Thr  | Cys | Leu   | ----- | Ala | phe   | lle | ys  | Pro   | Ala | lle | Asn   | Asn | lle | Se    | 150 |     |       |     |       |       |       |       |
|    | GG    | CA   | AG  | TT    | GG   | TT  | T     | AT    | TG  | CA    | AG  | GG  | TT    | T   | AT  | TG    | CA  | AG  | GG    | TT  | T   | AT    | TG  | CA    | AG    | TT    | 89504 |
|    | r     | Leu  | Met | lle   | Leu  | Gly | Leu   | Pro   | Cys | Thr   | Ala | leu | Leu   | Val | Ala | Glu   | Leu | Lys | Arg   | Cys | As  | 170   |     |       |       |       |       |
|    |       |      |     |       |      |     |       |       |     |       |     |     |       |     |     |       |     |     |       |     |     |       |     |       |       |       |       |
|    | TT    | AT   | T   | A     | A    | A   | T     | ----- | GC  | AG    | T   | T   | G     | G   | T   | G     | G   | T   | G     | C   | T   | A     | CT  | T     | A     | ----- | 89538 |
|    | p     | Al   | n   | Val   | Arg  | Val | Phe   | Lys   | Leu | Gly   | Leu | Phe | Ser   | Gly | Leu | T     | Trp | Thr | Leu   | Ala | Leu | Ph    | 190 |       |       |       |       |
|    |       |      |     |       |      |     |       |       |     |       |     |     |       |     |     |       |     |     |       |     |     |       |     |       |       |       |       |
|    | ----- | AT   | AC  | T     | G    | CC  | AC    | T     | G   | GC    | AT  | G   | A     | T   | T   | C     | T   | G   | C     | G   | T   | G     | AG  | CT    | T     | TT    | 89588 |
|    | e     | Cys  | T   | Trp   | lle  | Ser | Arg   | P     | Ala | Phe   | Cys | l   | Leu   | Leu | Ser | Ser   | Phe | H   | i     | S   | Phe | Pro   | Tyr | Le    | 210   |       |       |
|    |       |      |     |       |      |     |       |       |     |       |     |     |       |     |     |       |     |     |       |     |     |       |     |       |       |       |       |
|    | G     | AT   | T   | G     | CA   | AG  | CA    | GG    | A   | T     | T   | T   | G     | G   | T   | T     | G   | T   | G     | T   | G   | T     | G   | T     | T     | ----- | 89589 |
|    | u     | His  | Cys | Val   | Trp  | H   | i     | S     | lle | Leu   | lle | Cys | Leu   | Ala | Ser | Tyr   | Leu | Gly | Cys   | Val | Cys | Phe   | Al  | 230   |       |       |       |
|    |       |      |     |       |      |     |       |       |     |       |     |     |       |     |     |       |     |     |       |     |     |       |     |       |       |       |       |
|    | ----- | TT   | AT  | G     | CA   | AG  | CA    | GG    | A   | T     | T   | T   | G     | G   | T   | T     | G   | T   | G     | T   | G   | T     | T   | ----- | 89610 |       |       |
|    | a     | Tyr  | Phe | Asp   | Ala  | Ala | ----- | Ser   | Glu | lle   | Pro | Glu | Gln   | Gly | Pro | Val   | 244 |     |       |     |     |       |     |       |       |       |       |
|    |       |      |     |       |      |     |       |       |     |       |     |     |       |     |     |       |     |     |       |     |     |       |     |       |       |       |       |
|    | GC    | AT   | T   | AT    | T    | AG  | T     | AC    | CC  | CA    | AA  | T   | T     | G   | A   | T     | AT  | CA  | T     | CA  | AT  | CA    | AA  | T     | ----- | 89661 |       |
|    | l     | lle  | Arg | Phe   | Trp  | Pro | Ser   | Glu   | Lys | ----- | Trp | Ala | Phe   | lle | G   | 1     | 257 |     |       |     |     |       |     |       |       |       |       |
|    |       |      |     |       |      |     |       |       |     |       |     |     |       |     |     |       |     |     |       |     |     |       |     |       |       |       |       |
|    | TG    | AC   | GG  | CA    | AT   | G   | CT    | T     | CA  | AT    | T   | CA  | AA    | C   | AG  | C     | AG  | T   | G     | G   | T   | T     | G   | CA    | AT    | TT    | 89780 |
|    |       |      |     |       |      |     |       |       |     |       |     |     |       |     |     |       |     |     |       |     |     |       |     |       |       |       |       |

RESULT 13  
US-09-540-236-426  
; Sequence 426, Application US/09540236  
; Patent No. 6673910  
; GENERAL INFORMATION:

|                        |               |
|------------------------|---------------|
| Alignment Scores:      |               |
| Pred. No.:             | 0.447         |
| Score:                 | 97.00         |
| Percent Similarity:    | 38.68%        |
| Best Local Similarity: | 25.00%        |
| Query Match:           | 6.43%         |
| DB:                    | 4             |
|                        | Length:       |
|                        | Matches:      |
|                        | Conservative: |
|                        | Mismatches:   |
|                        | Indels:       |
|                        | Gaps:         |
|                        | 2304          |
|                        | 53            |
|                        | 29            |
|                        | 83            |
|                        | 47            |
|                        | 10            |

US-10-017-410-2 (1-275) x US-09-540-236-426 (1-2304)

|    |     |  |      |
|----|-----|--|------|
| Qy | 74  | ValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAsp   | 93   |
|    |     |  |      |
| Db | 936 | GTATCAATTGGCAGG---TATTTC-----                                  | 980  |
|    |     |  |      |
| Qy | 94  | GlutLeuAlaIle-LeuTrpValIleuMetCysAlaLeuAlaMetTrpPheProArgArgTy | 113  |
|    |     |  |      |
| Db | 981 | GCATGTCGCTATTTTACGGGTGATTATTCAGACACTTGAACTTTTATGACGGATAAATT    | 1040 |
|    |     |  |      |
| Qy | 113 | rLeuProLysIlePheArgAsn---AspArgGlyArgPheLysAlaValValCysValle   | 132  |
|    |     |  |      |

Db 1041 TGCAACACATTTATGCACATTTGCCAAGATGCCAAATTCGTATGATGTCATGGTGT 1100  
Qy 132 userAlaileThrThrCysLeu-----AlaPheileLysProAlaileAsnAnileSe 150  
Db 1101 GGCAAGTTGGTTTATCGTGTGTTTACAGGGTTTGATGTCCTCGGTGCGAATGTTTA 1160  
Qy 150 rLeuMetileLysLeuProCysThrAlaLeuValAlaLysLeuLysArgCysAs 170  
Db 1161 TTTATTAAAT-----GCAGCTGGCTGGTTTCGCTACTTA----- 1194  
Qy 170 pAsnValArgValPheLysLeuGlyLeuPheSerGlyLeuTrpThrLeuAlaLeuPh 190  
Db 1195 -----ATATGCCCATCGCATGATTTCTGTCTGGCTGGGTAGGCTTGTGTAT 1244  
Qy 190 eCysTrpIleSerAspGlnAlaPheCysGluLeuLeuSerPheHisPheProTyrLe 210  
Db 1245 GATTGGCTTGAC-----CCTTATGT 1265  
Qy 210 uHisCysValTrpHisleileLysLeuAlaSerTyrLeuGlyCysValCysPheAl 230  
Db 1266 G-----TTATGGCAACAGAGATTTGGTTGTCTATTTATTCGCCGTGTTTATGATCG 1319  
Qy 230 aTyrPheAspAlaAlaSerGluileProGlu-----GlnGlyProValileAr 246  
Db 1320 TTATGAGTCCAAATCAATATGATGATGAATGACACCAATCAAAATCAGGGCACTGTACG 1379  
Qy 246 gPheTrpProSerGluLys-----TrpAlaPheileGlyValPr 259  
Db 1380 GCGATGGCTCAATCAAAACAGGAGTGGCTTTCGAAATTTTGGCTATTTTATGATGCT 1439  
Qy 259 oTyrValSerLeuLeuCysAlaHisLysLysSer 270  
Db 1440 ACCTGTCTCTATCTATTATTGTCGAAGGTATCA 1473

## RESULT 14

US-08-956-171E-63/c  
; Sequence 63, Application US/08956171E  
; Patent No. 6593114  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; Gil H. Choi  
; Patrick S. Dillon  
; Craig A. Rosen  
; Steven C. Barash  
; Michael R. Fannon  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5256  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/956,171E  
; FILING DATE: 20-Oct-1997  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/009,861  
; FILING DATE: January 5, 1996  
; APPLICATION NUMBER: 08/781,986  
; FILING DATE: January 3, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mark J. Hymen  
; REGISTRATION NUMBER: 46,789  
; REFERENCE/DOCKET NUMBER: PB248P1  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (240) 314-1224  
; TELEFAX: (301) 309-8439  
; INFORMATION FOR SEQ ID NO: 63:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8155 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 63:  
US-08-956-171E-63  
Alignment Scores:  
Pred. No.: 2.91 Length: 8155  
Score: 97.00 Matches: 45  
Percent Similarity: 31.36% Conservative: 24  
Best Local Similarity: 20.45% Mismatches: 107  
Query Match: 6.42% Indels: 44  
DB: 4 Gaps: 8  
US-10-017-410-2 (1-275) x US-08-956-171E-63 (1-8155)  
Qy 38 SerAsnValLeuPhePheLeuProIleCysMetCysLeuPheArg---GlnTyr 56  
Db 5454 AGCAATTGCTCTTCGCAATATCTTTAGTGTGTTGCTGGCTGAATCGTTCATTTT 5395  
Qy 57 AlaThrCys-----PheAsnSerGlyIleTyrLeuileTrpThrLeuValVal 74  
Db 5394 AGCAATTGCTGCTTTTATAGCCGCTTCAACTTCGCAATTTGTATGTGCTGCTATTC 5335  
Qy 75 GlyIleGlySerValTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGlu 94  
Db 5334 GGCATCAGCTGTTGTTTTCAGTTTGAACCTTGTGTAGCAGCTGCTTTTCTTCAGT 5275  
Qy 95 -----LeuAlaileLeuTrpValLeuMetCys 103  
Db 5274 TGTTGAGCGGTTATTCATCAATTCGTTTCTTGAGCTGTACTTTATTCGCAATTCG 5215  
Qy 104 AlaLeuAlaMetTrpPheProArgArgTyrLeuProLysIlePheArgAsnArgGly 123  
Db 5214 TTGTTTGTGCTGGTTT-----AACATTTGCATCAGGTGTAATGGC 5173  
Qy 124 -----ArgPheLysAlaValValCysValLeuSerAlaileThrCysLeuAla 140  
Db 5172 TGCATTTGTAGCTTCATTTGTAGTTTTCATTAATCCATCATCTGTTGCTGAGCAGCAT 5113  
Qy 141 PheileLysProAlaileAsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla 160  
Db 5112 ATCTATATCAGCGTTGTCAGTAACCTACTGCTTGTATCCACTTGTCTTCGCTGCTGTG 5053  
Qy 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180  
Db 5052 TTCTTCAGTAGTCGAATCAATTCATTTGCTTCAATTCG----- 5017  
Qy 181 SerGlyLeuTrpThrLeuAlaLeuPheCysTrpIleSerAspGlnAlaPheCys--- 199  
Db 5016 -----TGTTTACGTTCACTTGTCTTGTGCGGATTTCCGCTTTTGCATC 4972  
Qy 200 GluLeuLeuSerSerPheHisPheProTyrLeuHisCys----- 212  
Db 4971 CGATTTCTTAGTTGGCAGCTTGAACTTGATTAATTCAGCAATACTATTGCTTTAGC 4912  
Qy 213 ValTrpHisleileLysLeuAlaSerTyrLeuGlyCysValCysPheAlaTyrPhe 232  
Db 4911 TGTTGTTACATCACTGTTGTATTTGCGATCAAGATTTGTTCTGCTTCTTCTTTT 4852  
RESULT 15  
US-08-781-986A-63/c  
; Sequence 63, Application US/08781986A  
; Patent No. 673248  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 21, 2004, 07:44:08 ; Search time 467 Seconds  
(without alignments)

3182.005 Million cell updates/sec

Title: US-10-017-410-2

Perfect score: 1512

Sequence: 1 MGAPHHWDLHLAGSEVDWC.....IGVPVSLCAHKKSPVKIT 275

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 3627888 seqs, 2701811610 residues

Total number of hits satisfying chosen parameters: 7255776

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool.p/US10017410/runat.19112004.133621.19052/app.query.fasta\_1.455  
-DB=Published Applications NA -QPM=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10017410 @CGN 1 1 723 @runat.19112004.133621.19052  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
20: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
21: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Match | Length | DB | ID                   | Description        |
|------------|--------|-------|--------|----|----------------------|--------------------|
| 1          | 1512   | 100.0 | 4175   | 13 | US-10-017-410-1      | Sequence 1, Appl   |
| 2          | 1433   | 94.8  | 4202   | 16 | US-10-188-832-21     | Sequence 21, Appl  |
| 3          | 1433   | 94.8  | 4212   | 15 | US-10-295-027-151    | Sequence 151, App  |
| 4          | 1404   | 92.9  | 828    | 13 | US-10-017-410-3      | Sequence 3, Appl   |
| 5          | 1373.5 | 90.8  | 822    | 15 | US-10-182-447-6      | Sequence 6, Appl   |
| 6          | 665    | 44.0  | 744    | 10 | US-09-945-527-48     | Sequence 48, Appl  |
| 7          | 549.5  | 36.3  | 792    | 15 | US-10-182-447-4      | Sequence 4, Appl   |
| 8          | 220.5  | 14.6  | 1194   | 16 | US-10-425-114-32295  | Sequence 32295, A  |
| 9          | 220.5  | 14.6  | 1471   | 18 | US-10-425-115-107231 | Sequence 107231, A |
| 10         | 213.5  | 14.1  | 1547   | 17 | US-10-767-701-12261  | Sequence 12261, A  |
| 11         | 212.5  | 14.1  | 1436   | 17 | US-10-437-963-42300  | Sequence 42300, A  |
| 12         | 209.5  | 13.9  | 801    | 15 | US-10-182-447-5      | Sequence 5, Appl   |
| 13         | 209.5  | 13.9  | 1063   | 16 | US-10-302-172-875    | Sequence 875, App  |
| 14         | 206.5  | 13.7  | 1429   | 18 | US-10-425-115-77897  | Sequence 77897, A  |
| 15         | 186    | 12.3  | 35425  | 15 | US-10-017-161-2429   | Sequence 2429, Ap  |
| 16         | 186    | 12.3  | 35425  | 15 | US-10-232-798-2069   | Sequence 2069, Ap  |
| 17         | 185    | 12.2  | 1357   | 16 | US-10-424-599-34549  | Sequence 34549, A  |
| 18         | 165    | 10.9  | 275    | 18 | US-10-425-115-111725 | Sequence 111725, A |
| 19         | 161    | 10.6  | 636    | 16 | US-10-264-237-615    | Sequence 615, App  |
| 20         | 127.5  | 8.4   | 2186   | 16 | US-10-425-114-35305  | Sequence 35305, A  |
| 21         | 127    | 8.4   | 3034   | 18 | US-10-425-115-77895  | Sequence 77895, A  |
| 22         | 123.5  | 8.2   | 1498   | 15 | US-10-017-161-1999   | Sequence 1999, Ap  |
| 23         | 123.5  | 8.2   | 1498   | 15 | US-10-292-798-1645   | Sequence 1645, Ap  |
| 24         | 119.5  | 7.9   | 568    | 17 | US-10-767-701-30658  | Sequence 30658, A  |
| c          | 115    | 7.6   | 284    | 15 | US-10-002-631C-17    | Sequence 17, Appl  |
| 26         | 112.5  | 7.4   | 248436 | 13 | US-10-087-192-2014   | Sequence 2014, Ap  |
| 27         | 112    | 7.4   | 1501   | 15 | US-10-017-161-2097   | Sequence 2097, Ap  |
| 28         | 112    | 7.4   | 1501   | 15 | US-10-232-798-1743   | Sequence 1743, Ap  |
| 29         | 111    | 7.3   | 1880   | 14 | US-10-198-846-12701  | Sequence 12701, A  |
| 30         | 110.5  | 7.3   | 744802 | 15 | US-10-292-798-1369   | Sequence 1369, Ap  |
| 31         | 109    | 7.2   | 1282   | 15 | US-10-017-161-2099   | Sequence 2099, Ap  |
| 32         | 109    | 7.2   | 1282   | 15 | US-10-292-798-1745   | Sequence 1745, Ap  |
| 33         | 109    | 7.2   | 2485   | 14 | US-10-198-846-13523  | Sequence 13523, A  |
| c          | 108.5  | 7.2   | 5842   | 11 | US-09-984-429-316    | Sequence 316, App  |
| 35         | 108    | 7.1   | 1984   | 15 | US-10-017-161-2133   | Sequence 2133, Ap  |
| 36         | 108    | 7.1   | 2614   | 15 | US-10-232-798-1779   | Sequence 1779, Ap  |
| 37         | 107.5  | 7.1   | 338702 | 13 | US-10-087-192-292    | Sequence 292, App  |
| c          | 106.5  | 7.0   | 63828  | 13 | US-10-087-192-388    | Sequence 388, App  |
| 39         | 106    | 7.0   | 2195   | 18 | US-10-425-115-92688  | Sequence 92688, A  |
| c          | 105.5  | 7.0   | 509    | 15 | US-10-029-386-3107   | Sequence 3107, Ap  |
| 41         | 105.5  | 7.0   | 439892 | 13 | US-10-087-192-454    | Sequence 454, App  |
| 42         | 105    | 6.9   | 1310   | 17 | US-10-437-963-72944  | Sequence 72944, A  |
| c          | 104.5  | 6.9   | 53714  | 16 | US-10-052-482-142    | Sequence 142, App  |
| c          | 104    | 6.9   | 2540   | 17 | US-10-437-963-6583   | Sequence 6583, Ap  |
| c          | 102    | 6.7   | 876    | 13 | US-10-001-857-17     | Sequence 17, Appl  |

#### ALIGNMENTS

##### RESULT 1

US-10-017-410-1  
; Sequence 1, Application US/10017410  
; Publication No. US20020115094A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnham, Peggy J  
; APPLICANT: Graveel, Carrie R  
; TITLE OF INVENTION: Polynucleotide Differentially Expressed in Liver Cancer  
; FILE REFERENCE: 960296.97401  
; CURRENT APPLICATION NUMBER: US/10/017,410  
; NUMBER OF SEQ ID NOS: 4  
; CURRENT FILING DATE: 2001-12-14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4175  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (35)...(859)  
US-10-017-410-1

```
Alignment Scores:
Pred. No.: 4,29e-157 Length: 4175
Score: 1512.00 Matches: 275
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-10-017-410-2 (1-275) x US-10-017-410-1 (1-4175)
QY 1 MetGlyAlaProHisTrpTyrAspHisLeuArgAlaGlySerSerGluValAspTrpCys 20
DB 35 ATGGGCGCCCGCAGCTGGTGGACACCTCGGGCTGGCAGTTCGGAGGTGGATTGGTGC 94
QY 21 GluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrAsnThrIleSerAsnVal 40
DB 95 GAGGACAACTACACTATCGTGCCTGCCATTGCCGAGTTCTACAACACGATCAGCAACGTC 154
QY 41 LeuPhePheIleLeuProProLleCysMetCysLeuPheArgGlnTyrAlaThrCysPhe 60
DB 155 TTGTTTTCATTTTACCTCCCATCTGCATGTGCTTGTTCGCCAGTAGTACGCAACGTGCTTC 214
QY 61 AsnSerGlyIleTyrLeuIleTyrThrLeuValValValGlyIleGlySerValTyr 80
DB 215 AACAGCGGCATCTACTTAATATGACCGCTCTAGTTGTAGTGGGATGGATCTGCTAC 274
QY 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTrpVal 100
DB 275 TTCCATGCAACGGCTAGTTTCTGGGTGAGTGTCTGATGCTTGAATGCACTTCTGTGGGT 334
QY 101 LeuMetCysAlaLeuAlaMetTyrPheProArgArgTyrLeuProLysIlePheArgAsn 120
DB 335 CTGATGTGTCTTGGCCATGTGCTTCCAGGAGGTATTACCAAGATCTTTCGGAAT 394
QY 121 AspArgGlyArgPheLeysAlaValValCysValLeuSerAlaIleThrCysLeuAla 140
DB 395 GACAGGGCAGGTTCAGGCGAGTGTGTGTCTGTCTGCTCAATACAGTGTCTGGCG 454
QY 141 PheIleLysProAlaIleAsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla 160
DB 455 TTTATCAAGCCCGCCATCAACAAATATTTCCCTGATGATTCTGGGACTTCCATGCACTGG 514
QY 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180
DB 515 CTGCTGTTCAGAGCTGAAGAGGTGCAATGTGGGTGTGTGTAAAGCTGGGCGCTCTTC 574
QY 181 SerGlyLeuTrpTrpThrLeuAlaLeuPheCysTrpIleSerAspGlnAlaPheCysGlu 200
DB 575 TCTGGCTCTGTGTGACTCTGGCTCTCTCTGCTGGATCAGCGACCAAGCTTCTGTGAG 634
QY 201 LeuLeuSerSerPheHisPheProTyrLeuHisCysValTrpHisIleLeuIleCysLeu 220
DB 635 CTGCTCTCTCTCTTCACTTCCCTACCTGCACTGTGTGGCATATTTCTCATCTGCCCT 694
QY 221 AlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAlaAlaSerGluIleProGlu 240
DB 695 GCCTTGTACCTGGCTGTGTGTCTGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTG 754
QY 241 GlnGlyProValIleArgPheTrpProSerGluLysTrpAlaPheIleGlyValProTyr 260
DB 755 CAAGGTCAGTCATCAGATTCGGCCCGCAGGAGAAATGGCTTTTATGTTGTCCTTAT 814
QY 261 ValSerLeuLeuCysAlaHisLysLysSerProValLysIleThr 275
DB 815 GTGTCCCTTCTGTGTGCCCCACAGAAAGTCCCGCAGTCAAGATCAG 859

RESULT 2
US-10-188-832-21
; Sequence 21, Application US/10188832
; Publication No. US20040076955A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
```

```
Db 617 TCGGGCTCTGGTGGACCTGGCCCTGTCTGTGGATCAGTGACCGAGCTTTCTGGAG 676
Qy 201 LeuLeuSerSerPheHisPheProTyrLeuHisCysValTyrPheHisLeuLeuLeuLeu 220
Db 677 CTGCTGTATCTTCACTTCCCTTACCTGACCTGCATGTGGTGGACATCTCTGCTT 736
Qy 221 AlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAlaAlaSerGluLeuProGlu 240
Db 737 GCTGCTACCTGGCTGTGTATGCTTGGCTTACTTGTATGCTGCTCAGAGATTCCTGAG 796
Qy 241 GlnGlyProValIleArgPheTyrProSerGluLysTyrPheAlaPheGlyValProTyr 260
Db 797 CAAGCCCTGTATCAAGATCTGCGCCCAATGAGAAATGGGCTTCATGGTCCCTAT 856
Qy 261 ValSerLeuLeuCysAlaHisLysLysSerProValLysIleThr 275
Db 857 GTGTCCCTCTGTGTGGCAACAGAAATCATCAGTCAAGATCAG 901

RESULT 3
US-10-295-027-151
; Sequence 151, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 151
; LENGTH: 4212
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-151

Alignment Scores:
Pred. No.: 2,64e-148 Length: 4212
Score: 1433.00 Matches: 256
Percent Similarity: 98.18% Conservative: 14
Best Local Similarity: 93.09% Mismatches: 5
```

```
Query Match: 94.78% Indels: 0
DB: 15 Gaps: 0
US-10-017-410-2 (1-275) x US-10-295-027-151 (1-4212)
Qy 1 MetGlyAlaProHisTyrTrpAspHisLeuArgHisSerSerGluValAspTyrCys 20
Db 87 ATGGGCGCCCGCAGCTGGTGGACACAGCTGAGCTGGTGGTGGTGGTGGTGGTGGTGG 146
Qy 21 GluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrAsnThrIleSerAsnVal 40
Db 147 GAGGACAACTACACCATCGTGGCTGCTATCGCCGAGTCTTACACACAGATCAGCAATGC 206
Qy 41 LeuPhePheIleLeuProProIleCysMetCysLeuPheArgGlnTyrAlaThrCysPhe 60
Db 207 TTATTTTTCATTTTACCGCCCATCTGCATGTGCTTGTTCGTAGTATGCAACATGCTTC 266
Qy 61 AsnSerGlyIleTyrLeuIleTyrThrLeuLeuValValGlyIleGlySerValTyr 80
Db 267 AACAGTGGCATCTACTTAATCTGGACTCTTTTGGTGTGTAGTGGAAATGGATCGTCTAC 326
Qy 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTyrVal 100
Db 327 TTCATGCAACCTTAGTTCTTGGGTGAGTGTGATGATGATGATGATGATGATGATGAT 386
Qy 101 LeuMetCysAlaLeuAlaMetTyrPheProArgArgTyrLeuProLysIlePheArgAsn 120
Db 387 CTGATGTGTGCTTTGGCCATGTGTTCCCGAAGGTATCTACCAAGATCTTTTCGGAAT 446
Qy 121 AspArgGlyArgPheLysAlaValValCysValLeuSerAlaIleThrThrCysLeuAla 140
Db 447 GACCGGGTAGGTTCAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 506
Qy 141 PheIleLysProAlaIleAsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla 160
Db 507 TTTGTCAAGCTGCCATCAACACATCTCTGTATGATGATGATGATGATGATGATGATGAT 566
Qy 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180
Db 567 CTGCTCATCGCAGAGCTAAAGAGGTGTGACAAACATGCGTGTGTGTTAAAGCTGGGCTCTTC 626
Qy 181 SerGlyLeuTyrTrpThrLeuAlaLeuPheCysTyrPheSerAspGlnAlaPheCysGlu 200
Db 627 TCGGGCTCTGGTGGACCTGGCCCTGTCTGTGGATCAGTGGACGAGCTTTCTGGCAG 686
Qy 201 LeuLeuSerSerPheHisPheProTyrLeuHisCysValTyrPheHisLeuLeuLeuLeu 220
Db 687 CTGCTGTATCTCTCAACTTCCCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 746
Qy 221 AlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAlaAlaSerGluIleProGlu 240
Db 747 GCTGCTACCTGGGCTGTGTATGCTTGGCTTACTTGTATGCTGCTGCTGCTGCTGCTGCTG 806
Qy 241 GlnGlyProValIleArgPheTyrProSerGluLysTyrPheAlaPheGlyValProTyr 260
Db 807 CAAGCCCTGTATCAAGTTCGCGCCCAATGAGAAATGGGCTTCATGGTGGTGGTGGTGGT 866
Qy 261 ValSerLeuLeuCysAlaHisLysLysSerProValLysIleThr 275
Db 867 GTGTCCCTCTGTGTGGCAACAGAAATCATCAGTCAAGATCAG 911

RESULT 4
US-10-017-410-3
; Sequence 3, Application US/10017410
; Publication No. US20020115094A1
; GENERAL INFORMATION:
; APPLICANT: Farnham, Peggy J
; APPLICANT: Graveel, Carrie R
; TITLE OF INVENTION: Polynucleotide Differentially Expressed in Liver Cancer
; FILE REFERENCE: 960296.97401
; CURRENT APPLICATION NUMBER: US/10/017,410
; CURRENT FILING DATE: 2001-12-14
; NUMBER OF SEQ ID NOS: 4
```

; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 828  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(825)  
US-10-017-410-3

Alignment Scores:  
Pred. No.: 4,15e-146 Length: 828  
Score: 1404.00 Matches: 251  
Percent Similarity: 96.73% Conservatives: 15  
Best Local Similarity: 91.27% Mismatches: 9  
Query Match: 92.86% Indels: 0  
DB: 13 Gaps: 0

US-10-017-410-2 (1-275) x US-10-017-410-3 (1-828)

```
QY 1 MetGlyAlaProHisTrpTrpAspHisLeuArgAlaGlySerSerGluValAspTrpCys 20
DB 1 ATGGGGCCCGCACCTGGTGGGACCCAGCTGCAGCTGGTAGCTCGGAGGTGGACTGGTGC 60
QY 21 GluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrAsnThrIleSerAsnVal 40
DB 61 GAGGACAACTACCACTGGCTGCTATCGCCGAGTTCTACCAACAGATCAGCAATGTC 120
QY 41 LeuPhePheIleLeuProProIleCysMetCysLeuPheArgGlnTyrAlaThrCysPhe 60
DB 121 TTATTTTTCATTTTACCGCCCATCTGATGTGCTTGTGATGAGTATGCAACATGCTTG 180
QY 61 AsnSerGlyIleTyrIleuileTrpThrLeuValValValGlyIleGlySerValTyr 80
DB 181 AACAGTGGACATCTACTTAATCTGACTCTTTGGTTGTAGTGGGAATTTGATCCGCTAC 240
QY 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTrpVal 100
DB 241 TTCCATTTTACCTTTAGTTTCTGGGTGATGCTTGTATGAACTTTCAGTCCCTTTGGGTT 300
QY 101 LeuMetCysAlaLeuAlaMetTrpPheProArgTyrLeuProLysIlePheArgAsn 120
DB 301 CTGATGTGTCTTTGGCCATGTGTTCCCGAGAGGTATCTACCAAGATCTTTTCGAAT 360
QY 121 AspArgGlyArgPheLysAlaValValCysValLeuSerAlaIleThrCysLeuAla 140
DB 361 GACAGGGGTAGGTTCAGAGTGGTGGTGCAGTGTCTCTGCGGTATACGAGTGCCTGGCA 420
QY 141 PheIleLysProAlaIleAsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla 160
DB 421 TTGTGCAAGCTGCCATCAACACATCTCTGATGACCTGGAGGTTCCTTGCATGCA 480
QY 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180
DB 481 CTGCTCATCGCAGAGCTAAGAGGTGTGACCAATCGTGTGTTTAAGCTGGGCCCTTTC 540
QY 181 SerGlyLeuTrpTrpThrLeuAlaLeuPheCysTyrIleSerAspGlnAlaPheCysGlu 200
DB 541 TCGGGCCCTCGTGGGACCTGGCCCTGTTCTGTGATGACCTGGAGGTTCCTTCTCGAG 600
QY 201 LeuLeuSerSerPheHisPheProTyrLeuHisCysValTrpHisIleLeuIleCysLeu 220
DB 601 CTGTGTGATCTTCAACTTCCCTACCTGCACTGATGTGGGACATCTCATCTGCTT 660
QY 221 AlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAlaIleAsnGluIleProGlu 240
DB 661 GCTGCCTACCTGGCTGTGTATGCTTTGCTACTTTGATGCTGCTCAGAGATTCCTGAG 720
QY 241 GlnGlyProValIleArgPheTrpProSerGluLysTrpAlaPheIleGlyValProTyr 260
DB 721 CAAGGCCCTGTCAATGATTTCTGCCCAATGAGAAATGGGCCCTTCATGTGTGCTCCCTAT 780
QY 261 ValSerLeuLeuCysAlaHisLysLysSerProValIysIleThr 275
```

DB 781 GTGTCCCTCTGTGTGCCAACAGAAATCATCATGATCAAGATCAG 825

## RESULT 5

US-10-182-447-6  
; Sequence 6, Application US/10182447  
; Publication No. US20030185814A1  
; GENERAL INFORMATION:  
; APPLICANT: HOFMANN, Kay  
; APPLICANT: RADT, Marcus  
; TITLE OF INVENTION: CERAMIDASE  
; FILE REFERENCE: P68055US0  
; CURRENT APPLICATION NUMBER: US/10/182,447  
; CURRENT FILING DATE: 2002-07-29  
; PRIOR APPLICATION NUMBER: PCT/EP01/00900  
; PRIOR FILING DATE: 2001-01-27  
; PRIOR APPLICATION NUMBER: DE 10003293.1  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: DE 10011392.3  
; PRIOR FILING DATE: 2000-03-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 822  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-182-447-6

Alignment Scores:  
Pred. No.: 1,01e-142 Length: 822  
Score: 1373.50 Matches: 246  
Percent Similarity: 94.91% Conservatives: 15  
Best Local Similarity: 89.45% Mismatches: 13  
Query Match: 90.84% Indels: 1  
DB: 15 Gaps: 1

US-10-017-410-2 (1-275) x US-10-182-447-6 (1-822)

```
QY 1 MetGlyAlaProHisTrpTrpAspHisLeuArgAlaGlySerSerGluValAspTrpCys 20
DB 1 ATGGGGCCCGCACCTGGTGGGACCCAGCTGCAGCTGGCAGCTCGGAGGTGGACTGGCGC 60
QY 21 GluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrAsnThrIleSerAsnVal 40
DB 61 GAGGACAACTACCACTGGCTGCTGCTGCGCGAGTTCTATACATGATCAGCAATGTC 120
QY 41 LeuPhePheIleLeuProProIleCysMetCysLeuPheArgGlnTyrAlaThrCysPhe 60
DB 121 TTATTTTTCATTTTACCGCCCATCTGATGTGCTTGTTCGTGATGATGCAACATGCTTC 180
QY 61 AsnSerGlyIleTyrIleuileTrpThrLeuValValValGlyIleGlySerValTyr 80
DB 181 AACAGCGGCATCTACTTAATCTGG---CTCTTGGTTGTAGCGGGAATTTGATCCGCTAC 237
QY 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTrpVal 100
DB 238 TTCATGCAACCTTCTAGTTTCTGGGTGATGCTTGTATGAACTTGCAGTCCCTTGGGTT 297
QY 101 LeuMetCysAlaLeuAlaMetTrpPheProArgTyrIleuProLysIlePheArgAsn 120
DB 298 CTGATGTGTCTTCGGTTCATGTGTTCCCGAGAGGTATCTACCAAGATCTTTTCGAAT 357
QY 121 AspArgGlyArgPheLysAlaValValCysValLeuSerAlaIleThrCysLeuAla 140
DB 358 GACCAGGGTAGGTTCAGAGTGGTGGTGTGCTGCTCTGCACTATGACGTGCTCGGCA 417
QY 141 PheIleLysProAlaIleAsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla 160
DB 418 TTGTGCAAGCTGCCATCAACACATCTCTCTGATGACCTGGAGATTCTTGGCTGCA 477
QY 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180
DB 478 CTGCTCATCAGAGAGCTAAAGAGGTGTGACACATCGTGTGTTTAAAGTGGGCTTCTTC 537
```



```

Db 489 GACCACATAAGGACCTTCGGACCTGATTGAGTCTCGTGTTTATGGGCTTTGC 548
QY 188 aLeuPheCysTrpIleSerAspGlnAlaPheCysGluLeuLeuSerSerPheHisPhePr 208
Db 549 TCTGACCACTGATGATGAGTGGTCTGCTTTGACGCTTCTGGCAGAGGATTCATTTCTT 608
QY 208 oTyrLeuHisCysValTrpHisIleLeuIleCysLeuAlaSerTyrLeuGlyCysValCy 228
Db 609 CTATCTGCACAGCATCTGGCATGTGCTCATCAGCATCACTTCCCTTATGGCATGTAC 668
QY 228 sPheAlaTyrPheAspAlaAlaSerGluIleProGluGinglyProValIleArgPheTr 248
Db 669 CATGGCTTGGTGGATGCCAATCATGATGATGATGATGATGATGATGATGATGATGATG 728
QY 248 pProSerGluLysTrpAlaPheIleGlyValProTyrValSerLeu 263
Db 729 GCTCGGGACAGTTGGCCC---GTGGGGGTGCTGCTTACGTGGAATC 771

RESULT 8
US-10-425-114-32295
; Sequence 32295, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53113)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 32295
; LENGTH: 1194
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73286S07_FLI
US-10-425-114-32295

Alignment Scores:
Pred. No.: 2,64e-14 Length: 1194
Score: 220.50 Matches: 76
Percent Similarity: 43.73% Conservatives: 39
Best Local Similarity: 28.90% Mismatches: 121
Query Match: 14.58% Indels: 27
DB: 16 Gaps: 9

US-10-017-410-2 (1-275) x US-10-425-114-32295 (1-1194)
QY 14 SerSerGluValAspTrpCysGluAspAsnTyrThrIleValProAlaIleAlaGluPhe 33
Db 117 ACATCAACGACTGAGTGTGGGAGAGATTTATGCACACTCGTCATATATCGCAGAAATTC 176
QY 34 TyrAsnThrIleSerAsnValLeuPhePheIleLeuLeuProIleCysMetCysLeuPhe 53
Db 177 TACAATACCATCTCTAATGTCCCATGCTTTTGGCCTTATT-----GGATTAGTG 230
QY 54 ArgGlnTyrAlaThrCysPheAsnSerGlyIleTyrLeuIleTrpThrLeuLeuValVal 73
Db 231 AATGCTTCTGCAAGTTTGGAGAACATTTAGTGTCTGACATATCCATATGATA 290
QY 74 ValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAsp 93
Db 291 CTTGCTATTGGGAGTATGATTTTCCATGCCACCTTCCCACTTCTCTACACAGAGTGTAT 350
QY 94 GluLeuAlaIleLeuTrpValLeuMetCysAlaLeuAlaMet-----Trp 108
Db 351 GAGACTCTATGTTGGGAGATCTCTCTCTACATGATGTCTCTCTATTCATTCACGGGACTGG 410

```

```

QY 109 PheProArgArgTyrLeuProLysIlePheArgAsnAspArgGlyArgPheLysAlaVal 128
Db 411 CACTACAGGAGACAAATGCCA---ACTTTCCTGTTCTCTGTATGCTGCTGCTTGGCGTA 467
QY 129 ValCysValLeuSerAlaIleThrThrCysLeuAlaPheIleLysProAlaIleAsnAsn 148
Db 468 GTACATTTCTTTGCCCGGTTCCAGTCGTATTTC-----AAGCTGCATTAC 512
QY 149 IleSerLeuMetIleLeuGlyLeuProCysThrAlaLeuLeuValAlaGluLeuLysArg 168
Db 513 GTTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 572
QY 169 CysAspAsnValArgValPheLysLeuGlyLeuPheSerGlyLeuTrpThrLeuAla 188
Db 573 GTGGGTGCGAAGCGGCTCGCAAACTGTGGTCT-----CTTACATGACCTTGGG 623
QY 189 LeuPheCysTrpIleSerAspGlnAlaPheCysGluLeuLeuSerSerPheHisPhe--- 207
Db 624 ACCCTCTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 683
QY 208 ProTyrLeuHisCysValTrpHisIleLeuIleCysLeuAlaSerTyrLeuGlyCysVal 227
Db 684 CGCAGGGCAGCGCGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 743
QY 228 CysPheAlaTyrPheAspAlaAlaSerGluIleProGluGlnGlyProValIleArgPhe 247
Db 744 ---TTCTGATGTTTGGCGGGCTCAGCAG-----CGCGGG 776
QY 248 TrpProSerGluLysTrpAlaPheIleGly---ValProTyrValSerLeuLeuCysAla 266
Db 777 TGGGAGCGCGGATCAGCACCTCTCTGATTTCTGCGGATCTTCCGATTCGATTCAGGTC 836
QY 267 HisLysLys 269
Db 837 GAGAAGAGG 845

RESULT 9
US-10-425-115-107231
; Sequence 107231, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 107231
; LENGTH: 1471
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_29295C.1
US-10-425-115-107231

Alignment Scores:
Pred. No.: 3.57e-14 Length: 1471
Score: 220.50 Matches: 76
Percent Similarity: 43.73% Conservatives: 39
Best Local Similarity: 28.90% Mismatches: 121
Query Match: 14.58% Indels: 27
DB: 18 Gaps: 9

US-10-017-410-2 (1-275) x US-10-425-115-107231 (1-1471)
QY 14 SerSerGluValAspTrpCysGluAspAsnTyrThrIleValProAlaIleAlaGluPhe 33
Db 303 ACATCAACGACTGAGTGTGGGAGAGATTTATGCACACTCGTCATATATCGCAGAAATTC 362

```

QY 34 TyrAsnThrIleSerAsnValLeuPhePheIleLeuProProIleCysMetCysLeuPhe 53  
Db 363 TACAATACCATCTCTTAATGTCCCATGCGCTCTTTGGCAGCACTAAT-----GGATTAGT 416  
QY 54 ArgGlnTyrAlaThrCysPheAsnSerGlyIleTyrLeuIleTyrThrLeuLeuValVal 73  
Db 417 AATGCTTCTGCCAAGGTTTGGAGAAAGATTTAGTGTCTGACATATCCAAATGATA 476  
QY 74 ValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAsp 93  
Db 477 CTGTGTAATGGAGTATGATTTTCCATGCCACCTTTGCAACTCTCTCTCAACAGAGTGT 536  
QY 94 GluLeuAlaIleLeuTyrValLeuMetCysAlaLeuAlaMet-----Tyr 108  
Db 537 GAGACTCTCTGTTGGAGATCTCTCTACATGATGCTCTCTATTCACCGAGCTGG 596  
QY 109 PheProArgGlyTyrLeuProLysIlePheArgAsnAspArgGlyArgPheLysAlaVal 128  
Db 597 CACTACAGAGCACAATGCCA---ACTTTCCTGTCTCCGTATGGTGTGCTTGGCGTA 653  
QY 129 ValCysValLeuSerAlaIleThrThrCysLeuAlaPheIleLysProAlaIleAsnAsn 148  
Db 654 GTACATTTCTTGGCCGGTTCACAGTCGTATTTC-----AAGCTGCATTAC 698  
QY 149 IleSerLeuMetIleLeuGlyLeuProCysThrAlaLeuLeuValAlaGlnLeuLysArg 168  
Db 699 GTTGGCTCTCTGCTCTGATCCCGCGATGTACAAAGTACTACATACAGACGAAAGAC 758  
QY 169 CysAspAsnValArgValPheLysLeuGlyLeuPheSerGlyLeuTyrTyrThrLeuAla 188  
Db 759 GTGGTGGCAAGCGCTCGCAAACTGTGGGTC-----CTTACACTGACCTTGGG 809  
QY 189 LeuPheCysTrpIleSerAspGlnAlaPheCysGlyLeuLeuSerSerPheHisPhe--- 207  
Db 810 ACCCTCTGCTGCTGGTGTATCGCTCTCTGCAAGAGCTTTGGCAATTGGTAGCTCAAC 869  
QY 208 ProTyrLeuHisCysValTyrPheIleLeuLeuCysLeuAlaSerTyrLeuGlyCysVal 227  
Db 870 CCGCAGGGCGCAGCGGTGGTGACGCTGTATGGGCTCAACTCGTACTATGCAAAACACG 929  
QY 228 CysPheAlaTyrPheAspAlaAlaSerGluIleProGluGlnGlyProValIleArgPhe 247  
Db 930 ---TTCCTGATGTTCTGGCGGCTCAGCAG-----CGCGGG 962  
QY 248 TrpProSerGluLysTrpAlaPheIleGly---ValProTyrValSerLeuLeuCysAla 266  
Db 963 TGGAGCGCGGATCACGCACCTCTCTGGATTTCTTGGCGTATGTCAGGTCACGAAACCA 1022  
QY 267 HisLysLys 269  
Db 1023 GAGAAGAGG 1031

## RESULT 10

US-10-767-701-12261  
; Sequence 12261, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 12261  
; LENGTH: 1547  
; TYPE: DNA  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS34423\_1

## US-10-767-701-12261

Alignment Scores:  
Pred. No.: 2,31e-13 Length: 1547  
Score: 213.50 Matches: 79  
Percent Similarity: 43.56% Conservative: 36  
Best Local Similarity: 29.92% Mismatches: 120  
Query Match: 14.12% Indels: 29  
DB: 17 Gaps: 10

US-10-017-410-2 (1-275) x US-10-767-701-12261 (1-1547)

QY 14 SerSerGluValAspTrpCysGluAspAsnTyrThrIleValProAlaIleAlaGluPhe 33  
Db 613 ACGTCAACAGCTAGTGTGGGAGAGAAATTATGCACACTCATCATATATGCGAATTC 672  
QY 34 TyrAsnThrIleSerAsnValLeuPhePheIleLeuProProIleCysMetCys---Leu 52  
Db 673 TACAATACCATCTCTAATGTCCATGCGTCTCTTTGGCGGTATTGGATTAGTGAATGCT 732  
QY 53 PheArgGlnTyrAlaThrCysPheAsnSerGlyIleTyrLeuIleTyrThrLeuLeuVal 72  
Db 733 TTCGCCAA-----GGTTTGGAGAAACGGTTCAGTGTCTTCACGTATCCAATATG 783  
QY 73 ValValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeu 92  
Db 784 ATACTTGTCTATTGGGAGTATGATTTTCCATGCCACCTTGCACACTTCTCTCACAACAGAGT 843  
QY 93 AspGluLeuAlaIleLeuTyrValLeuMetCysAlaLeuAlaMet----- 107  
Db 844 GATGAGACTCCGATGGTCTGGAGATCTCTTATATATGTATGTCTTTATTCACCGGAC 903  
QY 108 TrpPheProArgArgTyrLeuProLysIlePheArgAsnAspArgGlyArgPheLysAla 127  
Db 904 TGGCACTACAGAGCACAATGCCA---ACTTTCCTTTTCTTATATATGCTGCTGCTTTGCA 960  
QY 128 ValValCysValLeuSerAlaIleThrThrCysLeuAlaPheIleLysProAlaIleAsn 147  
Db 961 GTAGTTCAATTTCTTGGCCGGTTCACAGTTGTATTTC-----AAGTTGCAT 1005  
QY 148 AsnIleSerLeuMetIleLeuGlyLeuProCysThrAlaLeuLeuValAlaGluLeuLys 167  
Db 1006 TACATTGTCTGTGCTCTCTGTCATCCCGGATGTACAAAGTACTACATGACGAAA 1065  
QY 168 ArgCysAspAsnValArgValPheLysLeuGlyLeuPheSerGlyLeuTyrTrpThrLeu 187  
Db 1066 GACATGGCGGCGAAGCGCTTGCAGAGCTTTGGGT-----CTTACATTAACCTT 1116  
QY 188 AlaLeuPheCysTrpIleSerAspGlnAlaPheCysGluLeuLeuSerSerPheHisPhe 207  
Db 1117 GGGACTCTCTGCTGCTAGTTGATGCGATCTCTGTAAAGAGCTTTCATATGTTATGTC 1176  
QY 208 ---ProTyrLeuHisCysValTyrPheIleLeuIleCysLeuAlaSerTyrLeuGlyCys 226  
Db 1177 AACCCGACGCGGCGACGATGCTGGCATGTGCTTATGGGCTCAACTCATATGTCGAAAC 1236  
QY 227 ValCysPheAlaTyrPheAspAlaAlaSerGluIleProGluGlnGlyProValIleArg 246  
Db 1237 ACA---TTCCTAATGTTTGGCGGCTCAGCAA-----CGT 1269  
QY 247 PheTrpProSerGluLysTrpAlaPheIleGly---ValProTyrValSerLeuLeuCys 265  
Db 1270 GGGTGGGAGCGCGGATCACACACTTCTTGGATTCTTGGCTTACGTCAGGTCGAGAA 1329  
QY 266 AlaHisLysLys 269  
Db 1330 CCAGAAAGAGG 1341

## RESULT 11

US-10-437-963-42300  
; Sequence 42300, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Bing  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
SEQUENCE ID: 38-21(53221)B  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 42300  
LENGTH: 1436  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_45566C.1  
US-10-437-963-42300

Alignment Scores:  
Pred. No.: 2,68e-13 Length: 1436  
Score: 212.50 Matches: 73  
Percent Similarity: 42.70% Conservative: 41  
Best Local Similarity: 27.34% Mismatches: 118  
Query Match: 14.05% Indels: 35  
DB: 17 Gaps: 10

US-10-017-410-2 (1-275) x US-10-437-963-42300 (1-1436)

Qy 14 SerSerGluValAspTrpCysGluAspAsnTyrThrIleValProAlaIleAlaGluPhe 33  
Db 260 ACTTCAACAACTGAGCTCTCGAGGAAATATGACACTCGTCATATATGCGAGATTC 319  
Qy 34 TyrAsnThrIleSerAsnValLeuPhePheIleLeuProIleCysMetCysLeuPhe 53  
Db 320 TACAATGACTGTTCTAATGTT-----CGATGTGTTCTTCTGGCACTT 361  
Qy 54 ArgGlnTyrAlaThrCysPheAsnSerGlyIleTyrLeuIleTrpThrLeuLeu----- 71  
Db 362 GTTGGACTTGTGAATGCTCTTCGCAAGGTTTGTGAGAAACGATTTCAGTGTCTGCACATA 421  
Qy 72 -----ValValValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLeuGly 89  
Db 422 TCCAATATGATGACTTGTCTCGGAGTATGATCTTCCATGCCACCTTACAGCAGTTTAA 481  
Qy 90 GlnMetLeuAspGluLeuAlaIleLeuTrpValLeuMetCysAlaLeuAlaMet----- 107  
Db 482 CAGCAGAGCATGAGACTCCCAATGGTGTGGAGATTCCTATATCTTTATGACTTTAT 541  
Qy 108 -----TrpPheProArgArgTyrLeuProLysIlePheArgAsnAspArgGlyArg 124  
Db 542 TCACCAGACTGGCATATCCGAGCAGCATGCTACTTCTCTT-----TTCCCTA 589  
Qy 125 PheLysAlaValValCysValLeuSerAlaIleThrThrCysLeuAlaPheIleLysPro 144  
Db 590 TAGCGTCTGCTTTGTCAGTAGTCAATTCCTGTGCGATTCCTCAAGTGGTATTCAAG--- 646  
Qy 145 AlaIleAsnIleSerLeuMetIleLeuGlyLeuProCysThrAlaLeuLeuValAla 164  
Db 647 ---TTGCATTACGTTGGCTCTGCTCTATGATCCAGGATGTACAGTACTACATA 703  
Qy 165 GluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPheSerGlyLeuTrp 184  
Db 704 CAAACTAAAGACATGGCTGCCAAGCTAGCAAGCTGTGGTT-----CTTACG 754  
Qy 185 TrpThrLeuAlaLeuPheCysTrpIleSerAspGlnAlaPheCysGluLeuLeuSerSer 204  
Db 755 TTGAGCTTGGGCACTCTTTGCTGGCTATTGATCGAATGTTCTGTAAGAGCTTTTCACAT 814  
Qy 205 PheHisPhe---ProTyrLeuHisCysValTrpHisIleLeuIleCysLeuAlaSerTyr 223

Db 815 TGGTACGTCAACCCACAGGGCAGCATGTTGTCATATTCTCATGGGATTTAATCATCAT 874  
Qy 224 LeuGlyCysValCysPheAlaTyrPheAspAlaAlaSerGluIleProGluGlnGlyPro 243  
Db 875 TTTGCAACACA---TTCTTAATGTTTTCGCGAGCTCAACAG----- 913  
Qy 244 ValIleArgPheTrp---ProSerGluLysTrpAlaPheIleGlyValProTyrValSer 262  
Db 914 -----CGTGGGTGGAGCCCAAAATACCCACTTTTCGGGTCTTTCGCTTATGTCAAG 967  
Qy 263 LeuLeuCysAlaHisLysLys 269  
Db 968 ATTCAAGAACCCCAAAAGAGG 988

## RESULT 12

US-10-182-447-5  
Sequence 5, Application US/10182447  
Publication No. US20030185814A1  
GENERAL INFORMATION:  
APPLICANT: HOFMANN, Kay  
APPLICANT: RADT, Marcus  
TITLE OF INVENTION: CERAMIDASE  
FILE REFERENCE: P68055USO  
CURRENT APPLICATION NUMBER: US/10/182,447  
CURRENT FILING DATE: 2002-07-29  
PRIOR APPLICATION NUMBER: PCT/EP01/00900  
PRIOR FILING DATE: 2001-01-27  
PRIOR APPLICATION NUMBER: DE 10003293.1  
PRIOR FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: DE 10011392.3  
PRIOR FILING DATE: 2000-03-09  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 801  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-182-447-5

Alignment Scores:  
Pred. No.: 2,47e-13 Length: 801  
Score: 209.50 Matches: 81  
Percent Similarity: 40.89% Conservative: 29  
Best Local Similarity: 30.11% Mismatches: 103  
Query Match: 13.86% Indels: 58  
DB: 15 Gaps: 7

US-10-017-410-2 (1-275) x US-10-182-447-5 (1-801)

Qy 14 SerSerGluValAspTrpCysGluAspAsnTyrThrIleValProAlaIleAlaGluPhe 33  
Db 43 ACTTCAACAACTGAGCTCTCGAGGAAATATGACACTCGGTGTCATCGCGAGTTC 102  
Qy 34 TyrAsnThrIleSerAsnValLeuPhePheIleLeuProIleCysMetCysLeuPhe 53  
Db 103 TGAATACATGAGTAAAC---CTGATCATGATATACCTCCAA---TGTTCGGTCAATTC 157  
Qy 54 ArgGlnTyrAlaThrCysPheAsnSerGlyIleTyrLeuIleTrpThrLeuValVal 73  
Db 158 AGAGTGTAGACGCGTCTGGAAGCGGTACATTTGCTCT-TATTTAGCACTCACAGTG 216  
Qy 74 ValGlyIleGlySerValTyrPheHisAlaThrIleSerPheLeuGlyGlnMetLeuAsp 93  
Db 217 GTAGGAATGGATCCTGCTGCTCCACATGACTCTGAAATATGAAATGCACTATTGGAT 276  
Qy 94 GluLeuAlaIleLeuTrpValLeuMetCysAlaLeuAlaMetTrpPheProArgArgTyr 113  
Db 277 GAATCCCAATGATATAC---AGCTGTTGCATATTTGTGTACTGTCATGTTGAATGTTTC 333  
Qy 114 LeuProLysIlePheArgAsnAspArgGlyArgPheLysAlaValValCysValLeuSer 133  
Db 334 AAGATCAAGAACTCAGTAAACTACCACTCTGCTTTTACCTTAGTCTTATTACAGTTAATA 393





```
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_171065C.1
US-10-425-115-77897

Alignment Scores:
Pred. No.: 1.23e-12 Length: 1429
Score: 206.50 Matches: 75
Percent Similarity: 43.68% Conservative: 39
Best Local Similarity: 28.74% Mismatches: 112
Query Match: 13.66% Indels: 35
DB: 18 Gaps: 10

US-10-017-410-2 (1-275) x US-10-425-115-77897 (1-1429)

QY 14 SerSerGluValAspTrpCysGluAspAsnThrIleValProAlaIleAlaGluPhe 33
DB 442 ACATCAACAAGTGGTGGGAGGAGATTTATGCACACTCATATATAGCAGATTC 501

QY 34 TyrAsnThrIleSerAsnValLeuPhePheIleLeuProPheCysMetCysLeuPhe 53
DB 502 TACAACAACAATTTCAATGTC-----CCATGTGTTCCTTTTGGCGCTT 543

QY 54 ArgGlnThrAlaThrCysPheAsnSerGlyIleTyrLeuIleTrpThrLeuLeu----- 71
DB 544 ATTGGATTAGTAATGCTTTCCGCAAGGTTTTCAGAGGAGATTTCAGTGTCTTCACATA 603

QY 72 -----ValValValGlyIleGlySerValTyrPheHisAlaThrLeuSerPheLeuGly 89
DB 604 TCCATATGATCTGCTATTTGGGAGTATGATTTTTCATGCCACCTTGCACCTTGTCTTA 663

QY 90 GlnMetLeuAspGluLeuAlaIleLeuTrpValLeuMetCysAlaLeuAlaMet----- 107
DB 664 CAACAGAGTGACGAGATCCCAATGCTCTGGGAGATCCCTCTTTATATGATGTCTTAT 723

QY 108 -----TrpPheProArgTyrLeuProLysIlePheArgAsnAspArgGlyArg 124
DB 724 TCGCGGAGTGGCATTTACAGAGGACAAATGCCA---ACCTTCCTTTTCTATATGGGCT 780

QY 125 PheLysAlaValValCysValSerAlaIleThrThrCysLeuAlaPheIleLysPro 144
DB 781 GCCTTTGCTGTAGTTTCATTTTCCCGCGGTTCCAATTCGTATTC-----825

QY 145 AlaIleAsnIleSerLeuMetIleLeuGlyLeuProCysThrAlaLeuValAla 164
DB 826 AAGCTGCATTACATTTGGCTCTGCTTCTCTGCAATCCCGGATGTTCAAGTATTACATA 885

QY 165 GluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPheSerGlyLeuTrp 184
DB 886 CAGAGAAAGACGTGGCTGGCAACCGCTCCGAAACTGTGGGT-----CTCCCG 936

QY 185 TrpThrLeuAlaLeuPheCysTrpIleSerAspGlnAlaPheCys---GluLeuLeuSer 203
DB 937 TTAACCTTTGGGACTCTCTGCTGCTAGTTCATGTCATTTTTCAAAAGCTTTTGAT 996

QY 204 SerPheHisPheProTyrLeuHisCysValTrpHisIleLeuLeuCysLeuAlaSerTyr 223
DB 997 TGGTACGTCAACCCCGGAGGACGATGGTGGCATGTGCTTATGGGCTCAATCATAC 1056

QY 224 LeuGlyCysValCysPheAlaTyrPheAspAlaAlaSerGluIleProGluGlnGlyPro 243
DB 1057 TATGCAACCA---TTCTTATGTTTGTGGGCTCACAG-----1095

QY 244 ValIleArgPheTrpProSerGluLysTrpAlaPheIleGly---ValProTyrValSer 262
DB 1096 -----CGCGGGTGGGAGGCAAAATCCCTCCCTTCTTGGATTTTTCCTTATGTCAG 1149

QY 263 Leu 263
DB 1150 GTC 1152

RESULT 15
```

```

Db      21352  ||| .....:
Qy      75  -----GlyIleGlySe 78
Db      21412  GGGTTTAGGAGTGGCGGACCCCACTGACCGCTGCCCTTGCCGCTGCAGGCTGTTC 21471
Qy      78  rVallTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLe 98
Db      21472  CATGTATTTCCACATGACGCTCAGCTTCCTGGGCCAGCTGCTGGACGAGATGCCATCCT 21531
Qy      98  uTrpValLeuMetCysAlaLeuAlaMetTrpPheProArgArgTyrLeuProLysIlePh 118
Db      21532  GTGGCTCCTGGCGAGTGGCTATAGCATATGGATGCCCGCTGCTATTCCCTCCTTCCT 21591
Qy      118  eArgAsnAspArg 122
Db      21592  TGGGGGGAACAGG 21604

```

Search completed: November 21, 2004, 09:54:07  
 Job time : 482 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 21, 2004, 06:57:39 ; Search time 2986 Seconds  
(without alignments)  
3355.970 Million cell updates/sec

Title: US-10-017-410-2

Perfect score: 1512

Sequence: 1 MGAPHWDLRAGSEVDWC.....IGVPVSLCAHKKSPVKIT 275

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool.p/US10017410/runat\_19112004\_133620\_19010/app\_query.fasta\_1.455  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0 -LIST=45  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -MODE=LOCAL  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10017410.cgn 1 1 5180 @runat\_19112004\_133620\_19010 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | ID         | Description        |
|------------|--------|-------------|--------|------------|--------------------|
| 1          | 1392   | 92.1        | 1022   | 5 BU511164 | BU511164 AGENCOURT |
| 2          | 1230   | 81.3        | 681    | 9 AY401891 | AY401891 Mus muscu |
| 3          | 1176   | 77.8        | 868    | 6 CA976684 | CA976684 AGENCOURT |
| 4          | 1176   | 77.8        | 1173   | 3 AK085306 | AK085306 Mus muscu |
| 5          | 1163   | 76.9        | 681    | 9 AY401890 | AY401890 Pan trogl |
| 6          | 1158.5 | 76.6        | 797    | 6 CA463294 | CA463294 AGENCOURT |
| 7          | 1157   | 76.5        | 681    | 9 AY401889 | AY401889 Homo sapi |
| 8          | 1110.5 | 73.4        | 1527   | 3 AF370405 | AF370405 Homo sapi |
| 9          | 1001   | 66.2        | 698    | 5 BU234223 | BU234223 603792086 |

|    |       |      |     |   |          |                    |
|----|-------|------|-----|---|----------|--------------------|
| 10 | 996   | 65.9 | 565 | 6 | CF115220 | CF115220 Shultzomi |
| 11 | 986   | 65.2 | 633 | 2 | BB660847 | BB660847           |
| 12 | 984   | 65.1 | 622 | 6 | CB723138 | UI-M-GHO-          |
| 13 | 964   | 63.8 | 781 | 5 | BU232554 | 603409105          |
| 14 | 920   | 60.8 | 548 | 6 | CF169369 | CF169369 B0812G07- |
| 15 | 908   | 60.1 | 648 | 4 | RJ060795 | RJ060795           |
| 16 | 904   | 59.8 | 937 | 5 | EX842808 | EX842808           |
| 17 | 896   | 59.3 | 847 | 5 | BU220481 | 603107516          |
| 18 | 855   | 56.5 | 678 | 5 | BU203269 | 604153603          |
| 19 | 834   | 55.2 | 498 | 4 | BI848265 | 470659 MA          |
| 20 | 773   | 51.1 | 432 | 6 | CF169808 | B0818D08-          |
| 21 | 766   | 50.7 | 449 | 5 | EX646596 | DKF25781B          |
| 22 | 715   | 47.3 | 454 | 6 | CB784796 | AMGNNUC.T          |
| 23 | 692   | 45.8 | 401 | 2 | BF554219 | UI-R-CO-H          |
| 24 | 691.5 | 45.7 | 666 | 5 | BM215750 | BM215750           |
| 25 | 689.5 | 45.6 | 725 | 5 | BM442024 | BM442024           |
| 26 | 684.5 | 45.3 | 803 | 7 | CF593785 | AGENCOURT          |
| 27 | 669   | 44.2 | 427 | 1 | AA900336 | UI-R-E0-C          |
| 28 | 663   | 43.8 | 434 | 6 | CB758986 | AMGNNUC.M          |
| 29 | 653.5 | 43.2 | 815 | 5 | EX623644 | EX623644           |
| 30 | 651   | 43.1 | 399 | 6 | CB706492 | AMGNNUC.M          |
| 31 | 644.5 | 42.6 | 741 | 5 | BM216493 | BM216493           |
| 32 | 629.5 | 42.6 | 480 | 6 | CB728612 | AMGNNUC.M          |
| 33 | 628.5 | 41.6 | 696 | 5 | BM351328 | BM351328           |
| 34 | 625   | 41.3 | 485 | 2 | BF549345 | UI-R-A0-a          |
| 35 | 592   | 39.2 | 423 | 6 | CA967276 | CA967276           |
| 36 | 587   | 38.8 | 646 | 6 | CA969385 | CCLX06a22          |
| 37 | 575.5 | 38.1 | 712 | 5 | EX620332 | EX620332           |
| 38 | 570.5 | 37.7 | 398 | 6 | CB707581 | AMGNNUC.M          |
| 39 | 566   | 37.4 | 847 | 7 | CN060612 | CN060612           |
| 40 | 563   | 37.2 | 771 | 7 | CN057662 | Salamande          |
| 41 | 562   | 37.2 | 634 | 1 | AL864302 | AL864302           |
| 42 | 558   | 36.9 | 723 | 4 | BM602328 | BM602328           |
| 43 | 552   | 36.5 | 357 | 5 | BY168309 | BY168309           |
| 44 | 547   | 36.2 | 685 | 4 | BJ062108 | BJ062108           |
| 45 | 541   | 35.8 | 365 | 5 | BY117228 | BY117228           |

#### ALIGNMENTS

RESULT 1  
BU511164  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BU511164 1022 bp mRNA linear EST 12-SEP-2002  
AGENCOURT 10107530 NIH MGC 134 Mus musculus cDNA clone  
IMAGE:6505924 5', mRNA sequence.

BU511164  
EST.  
Mus musculus (house mouse)

BU511164.1 GI:22817397

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. David Rowe

cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM14067 Row: j Column: 05

High quality sequence stop: 681.

Location/Qualifiers

1..1022

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:6505924"

#### FEATURES

source



```

|||||
301 TCCCTGATGATCTCGGACTTCATGCACTCGCTGCTGTCGAGCTGAAGAGTGT 360
QY
170 AspAenValAlpPheLysLeuGlyLeuPheSerGlyLeuTrpThrLeuAlaLeu 189
Db
361 GACAAATGTGCGTGTGTTAAAGCTGGGCTCTTCTCTGGCTCTGCTGGACTCTGGCTCTC 420
QY
190 PheCysTrpLleSerAspGlnAlaPheCysGluLeuLeuSerSerPheHisPheProTrp 209
Db
421 TTCCTGCTGGATCAGCGACCAAGCCCTCTGTGTGAGTGTCTCTCTCTTTCACATTCGCCCTAC 480
QY
210 LeuHisCysValTrpHisLleLeuLleCysLeuAlaSerTrpLeuGlyCysValCysPhe 229
Db
481 CTGCACTGTGTGCAATATCTCATCTGCCCTTGCTTGTACTGGGCTGTGTGTCTTC 540
QY
230 AlaTrpPheAspAlaSerGluLeuProGluGlnGlyProValLleArgPheTrpPro 249
Db
541 GCCTACTTTGATGCTGCGCTCAGAGATACCTGAGCAAGGTCCAGTCCATCAGATTCCTGGCCC 600
QY
250 SerGluLysTrpAlaPheLleGlyValProTrpValSerLeuLeuCysAlaHisLysLys 269
Db
601 AGCGAGAAATGGCTTTTATTGGTGTCCCTTAATGCTGTCTTGTCTCTGTGTGCCACAGAAG 660
QY
270 SerProValLysIleThr 275
Db
661 TCGCCAGTCAAGATCAG 678

```

RESULT 3  
CA976684  
LOCUS  
DEFINITION AGENCOURT 8877974 NCI CGAP Mam2 Mus musculus cDNA clone  
IMAGE:6437893 5', mRNA sequence.

ACCESSION CA976684  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: LLAM13955 row: 9 column: 14  
High quality sequence stop: 679.  
Location/Qualifiers  
1..868  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N-3"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6437893"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Mam2"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

FEATURES  
source

Score: 1176.00 Matches: 214  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 77.78% Indels: 0  
DB: 6 Gaps: 0

US-10-017-410-2 (1-275) x CA976684 (1-868)

QY 1 MetGlyAlaProHisTrpTrpAspHisLeuArgAlaGlySerSerGluValAspTrpCys 20  
Db 12 ATGGCGCGCCGCACTGGTGGAGCCACCTCGGGCTGGCAGTTCGAGGTGATTGGTGC 71  
QY 21 GluAspAsnTrpThrIleValProAlaLleAlaGluPheTrpAsnThrIleSerAsnVal 40  
Db 72 GAGGACAACTACATCTGTCCTGCCATTGCCAGTTCTACCAACGATCAGCAACGTC 131  
QY 41 LeuPhePheLleLeuProPheCysMetCysLeuPhePheArgGlnTrpAlaThrCysPhe 60  
Db 132 TTGTTTTTCATTTTACCTCCCATCTGCATGTGCTTGTCCGCCAGTACCAACGTCCTTC 191  
QY 61 AsnSerGlyLleTrpLeuLleTrpThrLeuLeuValValGlyIleGlySerValTrp 80  
Db 192 AACAGCGGCATCTACTTAATATGGAGCGCTCCTAGTTGTAGTGGGATTGGATCTGTCTAC 251  
QY 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaLleLeuTrpVal 100  
Db 252 TTCCATGCAACGCTGAGTTTCTTGGGTGAGTGTGATGAATTCGCAATTCGTGGGT 311  
QY 101 LeuMetCysAlaLeuAlaMetTrpPheProArgArgTrpLeuProLysIlePheArgAsn 120  
Db 312 CTGATGTGTGCTTTGGCCATGTGGTTCCAGAGAGTATTACCAAGATCTTCGGAAT 371  
QY 121 AspArgGlyArgPheLysAlaValValCysValLeuSerAlaLleThrThrCysLeuAla 140  
Db 372 GACAGGGCGAGGTTCAAGCGAGTGTGTGTCTGTCTGCAATTAACAACGTGTGGCG 431  
QY 141 PheLleLysProAlaLleAsnAsnLleSerLeuMetLleLeuGlyLeuProCysThrAla 160  
Db 432 TTATCAAGCCCGCCATCAACAATATTTCCCTGATGATTCTGGGACTTCATGCACTGCG 491  
QY 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180  
Db 492 CTGCTTGTGTGAGAGCTGAAGAGGTGACAAATGTGCGTGTGTTTAAAGCTGGGCTCTTC 551  
QY 181 SerGlyLeuTrpTrpThrLeuAlaLeuPheCysTrpLleSerAspGlnAlaPheCysGlu 200  
Db 552 TCTGGCCTCTGGTGACTCTGGCTCTCTTCTGCTGGATCAGCGACCAAGCCCTTCGTGTAG 611  
QY 201 LeuLeuSerSerPheHisPheProTrpLeuHisCysValTrp 214  
Db 612 CTGCTCTCTCTCTTTCCTTCCCTACCTGCACTGTGTGTGG 653

RESULT 4  
LOCUS  
DEFINITION

AK085306 1173 bp mRNA linear HTC 03-APR-2004  
Mus musculus 0 day neonate kidney cDNA, RIKEN full-length enriched  
library, clone:DG30008P07 product:similar to CANCER RELATED  
GENE-LIVER 1 [Mus musculus], full insert sequence.

ACCESSION AK085306  
VERSION AK085306.1 GI:26351558  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Alignment Scores:  
Pred. No.: 6.5e-111 Length: 868

Itch, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 20499374  
 MEDLINE  
 PUBMED  
 11042159

## REFERENCE

Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
 Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 MEDLINE  
 PUBMED  
 11076861

## REFERENCE

The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409, 685-690 (2001)  
 20499374

## REFERENCE

The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 20499374

## REFERENCE

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanganaki, T., Hara, A., Hashizume, W.,  
 Hayashida, K., Hayashida, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kageura, I., Kasukawa, T.,  
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
 Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
 Okazaki, Y., Saito, K., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
 Muramatsu, M. and Hayashizaki, Y.  
 Direct Submission

## REFERENCE

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,  
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.

## FEATURES

URL: http://genome.gsc.riken.jp/  
 Location/Qualifiers  
 1..1173  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM,DB:D630008P07"  
 /db\_xref="taxon:10090"  
 /clone="D630008P07"  
 /tissue\_type="kidney"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="0 day neonate"  
 72..731  
 /note="unnamed protein product; putative  
 similar to CANCER RELATED GENE-LIVER 1 [Mus musculus]"

## CDS

(SPTR|AAL40409, evidence: FASTV, 100%ID, 77.8%length,  
 match=642)"  
 /codon\_start=1  
 /protein\_id="BAC39416.1"  
 /db\_xref="GI:26351559"  
 /translation="MGAPHWMDHLRAGSSEVDNCTVTPVPAIEFYNTISNVLFPI  
 LPPTCMCLPRQYATCNSGIYLIWTLVVVIGSVYFATLSFGQMLDELAIIWVLM  
 CALAMPFPRYLKPIFENDGRKAVCVLSAITTCIAFLKPAIINNLSIMLIGLPCTA  
 LLVAELKRCNDNRVFKLGLFSLGSLWTUWALFCWISDQAFCELLSSHFHYLHCWSADR  
 G"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1e-110 Length: 1173  
 Score: 1176.00 Matches: 214  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 77.78% Indels: 0  
 DB: 3 Gaps: 0

US-10-017-410-2 (1-275) x AK085306 (1-1173)

Qy 1 MetGlyAlaProHisTrpTrpAspHisLeuArgAlaGlySerSerGluValAspTrpCys 20  
 Db 72 ATGGGCGCCCGCAGTCTGGTGGGACCACTGGCGGCTGGCAGTCTGGAGTGGTGTGC 131  
 Qy 21 GluAspAsnTrpThrIleValProAlaIleAlaGluPheTrpAsnThrIleSerAsnVal 40  
 Db 132 GAGGACAACTACACTATCGTGCCTCCATTCGCCAGTCTTACACACGATCAGCAACGTC 191  
 Qy 41 LeuPhePheIleLeuProIleCysMetCysLeuPheArgGlnTrpAlaThrCysPhe 60  
 Db 192 TTGTTTTCATTTTACCTCCCATCTGATGTGCTGTTCGCCAGTACGCAACGTCGTC 251  
 Qy 61 AsnSerGlyIleTrpLeuIleTrpThrIleLeuValValValGlyIleGlySerValTrp 80  
 Db 252 AACAGCGCATCTACTTAATATGACGCTCTAGTTGTAGTGGGATTTGGATCTGTCTAC 311  
 Qy 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleIleuTrpVal 100  
 Db 312 TTCCATGCAACGCTGAGTTTCTCGGTCAGATGCTTGTATGATCAATTCGCAATTCGTCG 371  
 Qy 101 LeuMetCysAlaLeuAlaMetTrpPheProArgTrpLeuProLysIlePheArgAsn 120  
 Db 372 CTGATGTGTGCTTGGCCATGTGGTTTCCAGAGAGATATTACCAAGATCTTTCGGAT 431  
 Qy 121 AspArgGlyArgPheLysAlaValValCysValLeuSerAlaIleThrThrCysLeuAla 140  
 Db 432 GACAGGGGAGGTTCAAGGCGAGTGTGTGTCTCTGTCTGCAATTACCAACGTCGTCGCG 491  
 Qy 141 PheIleLysProAlaIleAsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla 160  
 Db 492 TTTATCAAGCCCGCCATCAACAATTTTCCCTGATGATCTCGGACTTCCATGCACTCGG 551  
 Qy 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180  
 Db 552 CTGCTTGTTCAGAGCTGAAGAGGTGTGACATGTGGTGTGTGTTTAAGCTGGCTCTTC 611  
 Qy 181 SerGlyLeuTrpTrpThrLeuAlaLeuPheCysTrpIleSerAspGlnAlaPheCysGlu 200  
 Db 612 TCTGGCTCTGGTGGACTCTGGCTCTCTCTCTGCTGGATCAGCAGCAACGCTTCTGTGAG 671  
 Qy 201 LeuLeuSerSerPheHisPheProTrpIleHisCysValTrp 214  
 Db 672 CTGCTCTCTCTCTTACCTTCCCTACTGCACTGTGTGTGG 713

RESULT 5  
 AV401890  
 LOCUS  
 DEFINITION Pan troglodytes HCM1041 gene, VIRTUAL TRANSCRIPT, partial sequence,  
 genomic survey sequence.  
 AV401890  
 ACCESSION  
 VERSION AV401890.1 GI:39757876

AV401890 681 bp DNA linear GSS 12-DEC-2003



|   |   |   |     |
|---|---|---|-----|
| Db  | 421   | TTCCTGCGATCAGTACGCCAGCTTTCTTGCGAGCTGTCTGCATCCCTCAACTTCCCCTAC    | 480 |
| QY  | 210   | LeuHisCysValTrpHisIleLeuIleCysLeuAlaSerTyrrLeuGlyCysValCysPhe   | 229 |
| Db  | 481   | CTGCACTCATGTGGCACATCTCATCTGCTTCTGCTACCTACCTGGGCTGTGTATGCTTT     | 540 |
| QY  | 230   | AlaTyrrPheAspAlaAalaSerGlulileProGluGlnGlyProValIleArgPheTrpPro | 249 |
| Db  | 541   | GCCTACTTGTATGCTGCTCCTCAGATTCTCTGAGCAAGGCCCTGTCAATCAAGTTCTGGCCC  | 600 |
| QY  | 250   | SerGluLysTrpAlaPheIleGlyValProTyrrValSerLeuLeuCysAlaHisLysLys   | 269 |
| Db  | 601   | AATGAGAAATGGCTTCATTTGGTTCCTCATGTGCTCCCTCTCTGTGCAATAGAAA         | 660 |
| QY  | 270   | SerProValLysIleThr  | 275 |
| Db  | 661   | TCATCAGTCAAGATCACG  | 678 |
| RESULT 6  |   |   |     |
| LOCUS   | CA463294  |   |     |
| DEFINITION  | AGENCOURT 10691314 NIH MGC 169 Mus musculus cDNA clone  |   |     |
| IMAGE:  | 6770649 5', mRNA sequence.  |   |     |
| ACCESSION   | CA463294  |   |     |
| VERSION   | CA463294.1 GI:24919646  |   |     |
| KEYWORDS  | EST.  |   |     |
| SOURCE  | Mus musculus (house mouse)  |   |     |
| ORGANISM  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   |   |     |
| REFERENCE   | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  |   |     |
| AUTHORS   | 1 (bases 1 to 797)  |   |     |
| TITLE   | NIH-MGC http://mgc.nci.nih.gov/.  |   |     |
| JOURNAL   | National Institutes of Health, Mammalian Gene Collection (MGC)  |   |     |
| COMMENT   | Unpublished (1999)<br>Contact: Robert Strausberg, Ph.D.<br>Email: cgabbs-remail.nih.gov<br>Tissue Procurement: Dr. Jonathan Kuo, NIMH<br>cDNA Library Preparation: Michael Brownstein Laboratory<br>CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)<br>DNA Sequencing by: Agencourt Bioscience Corporation<br>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:<br>http://image.llnl.gov<br>Plate: LLC3088 row: p column: 08<br>High quality sequence stop: 536. |   |     |
| FEATURES  | Location/Qualifiers   |   |     |
| Source  | 1..797  |   |     |
| /organism="Mus musculus"  |   |   |     |
| /mol_type="mRNA"  |   |   |     |
| /db_xref="taxon:10090"  |   |   |     |
| /clone="IMAGE:6770649"  |   |   |     |
| /lab_host="DH10B (Ti-phage-resistant)"  |   |   |     |
| /clone_lib="NTH MGC 169"  |   |   |     |
| /note="Organ: Testicles; Vector: pDNR-LIB; Site 1: SfII (ggccattatggcc); Site 2: SfII (ggccctcgccc); cDNA made by oligo-dt priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:<br>5'-AAGCAGTGTATCAACGATGGCGCATACGCGCGG-3' and 5'-ATTCTAGAGCGGAGCGCGCATG-dt(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library." |   |   |     |
| ORIGIN  |   |   |     |
| Alignment Scores:   |   |   |     |
| Pred. No.:  | 3.71e-109   | Length:   | 797 |
| Score:  | 1158.50   | Matches:  | 227 |
| Percent Similarity:   | 89.31%  | Conservative:   | 7   |
| Best Local Similarity:  | 86.64%  | Mismatches:   | 21  |
| Query Match:  | 76.62%  | Indels:   | 3   |
| DB:   | 6   | Gaps:   |     |





PURMED  
COMMENT

12445392  
Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
Location/Qualifiers  
1. 698  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, Hisex"  
/db\_xref="taxon:9031"  
/clone="CHEST75K21"  
/dev\_stage="22"  
/lab\_host="DH10B"  
/clone\_lib="CSECHN24"  
/notes="Organ: heads; Vector: pBluescript II KS(+), Site\_1:  
ECORI; Site\_2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was blunted, ligated to NotI adapters, digested with  
ECORI, size-selected, and cloned into the NotI and ECORI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

FEATURES  
source

1. 698  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, Hisex"  
/db\_xref="taxon:9031"  
/clone="CHEST75K21"  
/dev\_stage="22"  
/lab\_host="DH10B"  
/clone\_lib="CSECHN24"  
/notes="Organ: heads; Vector: pBluescript II KS(+), Site\_1:  
ECORI; Site\_2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was blunted, ligated to NotI adapters, digested with  
ECORI, size-selected, and cloned into the NotI and ECORI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

ORIGIN

Alignment Scores:  
Pred. No.: 5.98e-93 Length: 698  
Score: 1001.00 Matches: 178  
Percent Similarity: 96.50% Conservativeness: 15  
Best Local Similarity: 89.00% Mismatches: 7  
Query Match: 66.20% Indels: 0  
DB: 5 Gaps: 0

US-10-017-410-2 (1-275) x BU234223 (1-698)

QY 76 IleGlySerValTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeu 95  
Db 1 ATTGGATCTCTTTTACTTTTCATGCGCCCTCAGCTTCTGGGTTCAGATGCTGGATGAGCTG 60  
QY 96 AlaileLeuTrpValLeuMetCysAlaLeuAlaMetTrpPheProArgArgTyrLeuPro 115  
Db 61 GCATTCCTGGGTCTCTATGTGCTCTTGGCATGGGTTCCTTAGGATATCTACCA 120  
QY 116 LysilePheArgAsnAspArgGlyArgPheLysAlaValValCysValLeuSerAlaile 135  
Db 121 AGAGTTTTTCGAAATGACAGACGCGGTTTAAAGCTGCTGTGTGCTGCTGTGGAGTT 180  
QY 136 ThrThrCysLeuAlaPheileLysProAlaileAsnAsnileSerLeuMetileLeuGly 155  
Db 181 ACTACCTGCTCTGCTTCATTAACCTGCCATCAACCACTCTCACTAAATGACTTGGGT 240  
QY 156 LeuProCysThrAlaLeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPhe 175  
Db 241 GTTCCTTGCACAGCTTTACTCATGCTGAGTTGAAGAGGTGTGAACACCTGGGTGTATC 300  
QY 176 LysLeuGlyLeuPheSerGlyLeuTrpTrpTrpTrpLeuAlaLeuPheCysTrpIleSerAsp 195  
Db 301 AAGCTTGCTCTGTTTTTTCAGGTCTTTGGTGATGCTAGCACTTTTCTGCTGGATCAGTGAC 360  
QY 196 GlnAlaPheCysGluLeuLeuSerPheHisPhePheProTrpTrpTrpHisCysValTrpHis 215  
Db 361 AAAGCTTTTGTGAGATCTGTGTATCACTTTAACTTCCCTTATTGCACTGTGTATGGCAC 420

QY 216 IleleuileCysleuAlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAlaAla 235  
Db 421 ATTTTGAATTTGCTTGGCTTACCTAGGCTGTGTGTTCCTTACTTCGATGCTGCC 480  
QY 236 SerGluileProGluGlnGlyProValileArgPheTrpProSerGluLysTrpAlaPhe 255  
Db 481 TCCAGATCCCTGAGCAGGCGCCGCTCATTAAGTTCTGGCCAGTGAAGATGGGCATTC 540  
QY 256 IleGlyValProTyrValSerLeuLeuCysAlaHisLysLysSerProValLysileThr 275  
Db 541 ATTGGGCTTCCCTACGTCACCTCTCTGTGCACACAAAGAAATCACCGGTGAAGATCACA 600

RESULT 10  
CF115220

LOCUS CF115220 565 bp mRNA linear EST 23-JUL-2003  
DEFINITION Shultzomica08471 Rat lung airway and parenchyma cDNA libraries  
Rattus norvegicus cDNA clone NP6159 5', mRNA sequence.

ACCESSION

CF115220  
VERSION CF115220.1 GI:33175919

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE

1 (bases 1 to 565)  
Shultz, M.A., Zhang, L., Gu, Y.-Z., Baker, G.L., Fannuchi, M.V.,  
Padua, A.M., Gurske, W.A., Morin, D., Penn, S.G., Jovanovich, S.B.,  
Plopper, C.G. and Buckpitt, A.R.

TITLE

Gene expression analysis in response to lung toxicants: I.  
Sequencing and microarray development

JOURNAL

Am. J. Respir. Cell Mol. Biol. 30 (3), 296-310 (2004)

COMMENT

Contact: Shultz MA  
Dept. of Molecular Biosciences, School of Veterinary Medicine  
University of California, Davis  
1311 Haring Hall, One Shields Avenue, Davis, CA 95616, USA  
Tel: 530 752 0793  
Fax: 530 752 4698  
Email: mashultz@ucdavis.edu  
Average Phred score is 20 or better. All poor quality data (Phred <  
20) and vector/linker sequence has been removed.  
High quality sequence stop: 565.

FEATURES  
source

1. 565  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="NP6159"  
/sex="male"  
/tissue\_type="airway or parenchyma"  
/dev\_stage="adult"  
/clone\_lib="Rat lung airway and parenchyma cDNA libraries"  
/notes="Organ: lung; Vector: pCEM-11zf(-); Site 1: Eco RI;  
Site 2: Not I; mRNA was isolated from microdissected rat  
lung airways and parenchyma tissues."

ORIGIN

Alignment Scores:  
Pred. No.: 1.45e-92 Length: 565  
Score: 996.00 Matches: 181  
Percent Similarity: 99.47% Conservativeness: 6  
Best Local Similarity: 96.28% Mismatches: 0  
Query Match: 65.87% Indels: 1  
DB: 6 Gaps: 0

US-10-017-410-2 (1-275) x CF115220 (1-565)

QY 74 ValGlyTleGlySerValTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAs 93  
Db 3 GTGGGAATCGAGTCTGTCTACTTCCATNGCAACTCTTAGTTCTCTGGTCAGATGCTTGA 62  
QY 93 pGluLeuAlaileLeuTrpValLeuMetCysAlaLeuAlaMetTrpPheProArgArgTy 113



Db 309 TTCCATGCAACGCTGAGTTTCTCTGGGTGAGATGCTTGATGAATGCTGCTGGGTT 368  
 QY 101 LeuMetCysAlaLeuAlaMetTTPheProArgArgTyrLeuProLysIlePheArgAsn 120  
 Db 369 CTGATGTGCTGCTTGGCCATGTGTGTTCCAGAGAGTATTACCAAGATCTTTCGGAAT 428  
 QY 121 AspArgGlyArgPheLysAlaValValCysValLeuSerAlaIleThrThrCysLeuAla 140  
 Db 429 GACAGGGCAGGTTCAAGGAGGTGTGTGCTGCTGCTCAATTACCAAGCTGTGGCG 488  
 QY 141 PheIleLysProAlaIleAsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla 160  
 Db 489 TTTATCAAGCCGCCATCAACATATTTCCCTGATGATTTCTGGACTTCCATGACTGCG 548  
 QY 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGly-LeuPh 180  
 Db 549 CTGCTTGTGTCAGAGCTGAAAGGTGTGACAATGTGGCTGTGTGTTAAATGGGCCCTTT 608  
 QY 180 eSerGlyLeuTTPheThrLeuAla 188  
 Db 609 TTTGGCCTGTGTGGACTCTGGCT 633

RESULT 12  
 CB723138  
 LOCUS  
 DEFINITION  
 UI-M-GH0-ceh-f-04-0-UI.r1 NIH BMAP\_GH0 Mus musculus cDNA clone  
 IMAGE:6839525 5', mRNA sequence.  
 CB723138  
 CB723138.1 GI:29780280  
 EST.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 622)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strauberg, Ph.D.  
 Email: [gsapbs-remail.nih.gov](mailto:gsapbs-remail.nih.gov)  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefl.html>  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)  
 Seq primer: pYX-5.

FEATURES  
 source  
 1..622  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6839525"  
 /tissue\_type="Whole brain"  
 /dev\_stage="1, 5, and 15 days newborn"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH BMAP GH0"  
 /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is CGAAGTGAAT. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National

Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
 program coordinator."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,9e-91 Length: 622  
 Score: 984.00 Matches: 180  
 Percent Similarity: 78.95% Conservative: 0  
 Best Local Similarity: 78.95% Mismatches: 2  
 Query Match: 65.08% Indels: 46  
 DB: 6 Gaps: 1

US-10-017-410-2 (1-275) x CB723138 (1-622)

QY 1 MetGlyAlaProHisTTPheProArgArgTyrLeuProLysIlePheArgAsnValAspTTPheCys 20  
 Db 76 ATGGGGCCCGCCGACCTGGTGGGACCACTGCGGGCTGGCAGTTCGGAGGTGGATTGGTGC 135  
 QY 21 GluAspAsnTyrThrIleValProAlaIleAlaGluPheTyrAsnThrIleSerAsnVal 40  
 Db 136 GAGGACAACTACACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 195  
 QY 41 LeuPhePheIleLeuProProIleCysMetCysLeuPheArgGlnTyrAlaThrCysPhe 60  
 Db 196 TTGTTTTTTCATTTTACCTCCCATCTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 255  
 QY 61 AsnSerGlyIleTyrLeuIleTTPheProArgArgTyrLeuProLysIlePheArgAsn 80  
 Db 256 AACAGGGCAGCTACTTAAATATGAGCTCTAGTGTAGTGGGATTTGATGCTGCTGCTGCT 315  
 QY 81 PheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLeuAlaIleLeuTTPheVal 100  
 Db 316 TTCCATGCAACGCTGAGTTTCTGGGTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 375  
 QY 101 LeuMetCysAlaLeuAlaMetTTPheProArgArgTyrLeuProLysIlePheArgAsn 120  
 Db 376 CTGATGTGCTTGTGGCCATGTGCTTCCAGAGAGTATTACCAAGATCTTTCGGAAT 435  
 QY 121 AspArgGlyArgPheLysAlaValValCysValLeuSerAlaIleThrThrCysLeuAla 140  
 Db 436 GAC-----AGGTGTGACAATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 438  
 QY 141 PheIleLysProAlaIleAsnAsnIleSerLeuMetIleLeuGlyLeuProCysThrAla 160  
 Db 438 -----AGGTGTGACAATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 438  
 QY 161 LeuLeuValAlaGluLeuLysArgCysAspAsnValArgValPheLysLeuGlyLeuPhe 180  
 Db 439 -----AGGTGTGACAATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 477  
 QY 181 SerGlyLeuTTPheThrLeuAlaLeuPheCysTTPheSerAspGlnAlaPheCysGlu 200  
 Db 478 TCTGGGCTCTGGTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAG 537  
 QY 201 LeuLeuSerSerPheHisPheProTyrLeuHisCysValTTPheIleLeuIleCysLeu 220  
 Db 538 CTGCT 597  
 QY 221 AlaSerTyrLeuGlyCysValCys 228  
 Db 598 GCTTCGTACTGGGGCTGTGTGTGC 621

RESULT 13  
 BU232554  
 LOCUS  
 DEFINITION  
 603409105F1 CSBQCHN24 Gallus gallus cDNA clone CHEST323j2 5', mRNA  
 sequence.  
 BU232554  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 BU232554 781 bp mRNA linear EST 26-NOV-2002  
 603409105F1 CSBQCHN24 Gallus gallus cDNA clone CHEST323j2 5', mRNA  
 sequence.  
 BU232554  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 BU232554.1 GI:25475352  
 Gallus gallus (chicken)  
 Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.

1 (bases 1 to 781)

Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.

A Comprehensive Collection of Chicken cDNAs

Curr. Biol. 12 (22), 1965-1969 (2002)

22335534

12445392

Contact: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology

(UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

Location/Qualifiers

FEATURES

source

Qy 175 heLysLeuGlyLeuPheSerGlyLeuTTPThrLeuAlaLeuPheCysTrpIleSerA 195  
Db 301 ACAAGCTGGTCTGTTTCAGGCTCTTGGTGGATGCTAGCAGCTTTCTGCTGGATCAGTG 360  
Qy 195 spGlnAlaPheCysGluLeuLeuSerSerPheHisPheProTyrLeuHisCysValTrpH 215  
Db 361 ACAAGCTTTTGTGAGATCTGTCATCATTTAACTTCCCTATTTGGACTGTGTATGGC 420  
Qy 215 isleLeuileCysLeuAlaSerTyrLeuGlyCysValCysPheAlaTyrPheAspAlaA 235  
Db 421 ACATTTGATTGCTTGGCGCTTACCTAGGCTGTGTCTGTTTCTGCTTACTTCGATGCTG 480  
Qy 235 laSerGluileProGluGlnGlyProValIleAArgPheTrpProSerGluLysTrpAlaP 255  
Db 481 CCTCGAGATCCCTGAGCAGGCCCCGTCATAAAGTCTTGCCCAAGTGAAGATGGGAT 540  
Qy 255 heLleGlyValProTyrValSerLeuLeuCysAlaHisLysLysSerProValLysIleT 275  
Db 541 TCATTTGGGTTCCCTACGTACCTCCTCTCTGTGCACACAGAATCACCAGTGAAGATCA 600  
Qy 275 hr 275  
Db 601 CA 602  
RESULT 14  
CF169369 548 bp mRNA linear EST 25-JUL-2003  
LOCUS B0812G07-5 NIA Mouse Newborn Kidney cDNA Library (long 1) Mus  
DEFINITION musculus cDNA clone NIA:B0812G07 IMAGE:30468654 5', mRNA sequence.  
ACCESSION CF169369  
VERSION CF169369.1 GI:33278918  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 548)  
AUTHORS Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.  
TITLE Construction of long-transcript enriched cDNA libraries from  
submicrogram amounts of total RNAs by a universal PCR amplification  
method  
JOURNAL Genome Res. 11 (9), 1553-1558 (2001)  
MEDLINE 21429098  
PubMed 11544199  
COMMENT Contact: Dawood B. Dudekula  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@lgsun.grc.nia.nih.gov  
Plate: B0812 row: G column: 07  
Seq primer: M13 Reverse  
High quality sequence stop: 548  
POLYA=No.

Qy 76 IleGlySerValTyrPheHisAlaThrLeuSerPheLeuGlyGlnMetLeuAspGluLe 95  
Db 1 ATTTGGATCTGTTTACTTTCATGGCCACCTCAGCTTCCTGGGTGAGATGCTGAGTAC 60  
Qy 95 uAlaIleLeuTrpValLeuMetCysAlaLeuAlaMetTrpPheProArgTyrTyrLeuPr 115  
Db 61 GGCTATTCTCTGGTGCCTTATGTGTGCTTGTGCCATGTGTTCCCTAGGAGATATCTACC 120  
Qy 115 olystlePheArgAsnAspArgGlyArgPheLysAlaValValCysValLeuSerAlaI 135  
Db 121 AAGAGTTTTTCGAATGACAGAGCGGTTTAAAGCTGCTGTGTGTCTGCTGAGT 180  
Qy 135 ePThrCysLeuAlaPheIleLysProAlaIleAsnLysSerLeuMetIleLeuG 155  
Db 181 TACTACCTGGCGCTTGCCTTCATTAACCTGCCATCAACAACTCTCATAATGACTCTGG 240  
Qy 155 lyLeuProCysThrAlaLeuLeuValAlaGluLeuLysArgCysAspAsnValArgValP 175  
Db 241 GTGTTTCCTTGCACAGCTTACTCATGTGTGAGTTCGAGAGGTGTGAAAACCTGCTGTGT 300

ORIGIN

Alignment Scores:

Pred. No.: 4.72e-89 Length: 781  
Score: 964.00 Matches: 177  
Percent Similarity: 95.05% Conservative: 15  
Best Local Similarity: 87.62% Mismatches: 8  
Query Match: 63.76% Indels: 2  
DB: 5 Gaps: 0

US-10-017-410-2 (1-275) x BU232554 (1-781)

FEATURES

source

Location/Qualifiers  
1.548  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="niaEST:B0812G07-5"  
/db\_xref="taxon:10090"  
/clone="NIA:B0812G07 IMAGE:30468654"  
/dev\_stage="Newborn Kidney"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse Newborn Kidney cDNA Library (Long 1)"  
/note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;  
Site 2: NotI; Mouse cDNA project by the Laboratory of  
Genetics, National Institute on Aging (NIA), Intramural  
Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).  
In brief, double-stranded cDNAs were synthesized with an  
Oligo(dT) primer [Invitrogen:  
5'-pGACTAGTTCATGATCGGCGGCCCTTTTCTTTT-3'] from  
26 ug of total RNA, treated with 14 DNA polymerase, and







***This Page Blank (uspto)***



construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES

Location/Qualifiers

1..2852

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CTONG202638"

/tissue type="tongue, tumor tissue"

/clone\_lib="CTONG2"

/note="cloning vector: pME18SFL3"

ORIGIN

Query Match 98.1%; Score 812; DB 9; Length 2852;

Best Local Similarity 98.8%; Pred. No. 8.9e-210;

Matches 818; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ATGGGCGCCCGCAGCTGGTGGGACGAGCTGAGGCTGGAGGTGGAGCTGGTGC 60

DB 159 ATGGGCGCCCGCAGCTGGTGGGACGAGCTGAGGCTGGAGGTGGAGCTGGTGC 218

QY 61 GAGGACAACTACACCATCGCTGCTATCGCCGAGTTCTACACACGATCAGCAATGTC 120

DB 219 GAGGACAACTACACCATCGCTGCTATCGCCGAGTTCTACACACGATCAGCAATGTC 278

QY 121 TTAATTTTCAATTTTACCGCCCATCTGCATGTGCTTTTGTGATGAGTATGCAATGCTTG 180

DB 279 TTAATTTTCAATTTTACCGCCCATCTGCATGTGCTTTTGTGATGAGTATGCAATGCTTG 338

QY 181 AACAGTGCATCTACTTAATCTGGACTCTTTTGGTGTAGTGGAAATGGATCGCTCTAC 240

DB 339 AACAGTGCATCTACTTAATCTGGACTCTTTTGGTGTAGTGGAAATGGATCGCTCTAC 398

QY 241 TTCCATTTTACCCCTTAGTCTTGGGTGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 300

DB 399 TTCCATTTTACCCCTTAGTCTTGGGTGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 458

QY 301 CTGATGTGTGCTTTGGGCAATGTTGCCAGAAAGGTATCTACCAAGATCTTTTGGGAT 360

DB 459 CTGATGTGTGCTTTGGGCAATGTTGCCAGAAAGGTATCTACCAAGATCTTTTGGGAT 518

QY 361 GACAGGGTGTGCTCAAGGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420

DB 519 GACAGGGTGTGCTCAAGGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 578

QY 421 TTTGTCAGGCTGCTGCTCAACCAATCTCTCTGATGACCCCTGGGAGTTTCTTTGCACTGCA 480

DB 579 TTTGTCAGGCTGCTGCTCAACCAATCTCTCTGATGACCCCTGGGAGTTTCTTTGCACTGCA 638

QY 481 CTGCTCATTCGAGAGCTAAAGAGGTGTGCAACATGCGTGTGTTTAAAGTGGGCTCTTTC 540

DB 639 CTGCTCATTCGAGAGCTAAAGAGGTGTGCAACATGCGTGTGTTTAAAGTGGGCTCTTTC 698

QY 541 TCGGGCTCTGCTGGGCACTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600

DB 699 TCGGGCTCTGCTGGGCACTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 758

QY 601 CTGCTGTCTCATCTTCAACTTCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660

DB 759 CTGCTGTCTCATCTTCAACTTCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 818

QY 661 GCTGCTACCTGGGCTGTGATGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720

DB 819 GCTGCTACCTGGGCTGTGATGCTTTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 878

QY 721 CAAGGCTCTGCTCAAGTCTTGGCCCAATGAGAAATGGGCTTCAATGGTGTGCCCTAT 780

DB 879 CAAGGCTCTGCTCAAGTCTTGGCCCAATGAGAAATGGGCTTCAATGGTGTGCCCTAT 938

QY 781 GTGTCCTCTGCTGTGGCAACAAGAAATCATCATGATCAAGATCAAGTGA 828

339 AACAGTGGCATCTACTTAATCTGGACCTCTTTTGGTGTAGTGGAAATGGATCGGTCTAC 398

241 TTCCATTTTACCCCTTAGTCTTGGGTGAGTGTGATGAACTTGCAAGTCTTTGGGTT 300

399 TTCCATGCAACCTTAGTCTTGGGTGAGTGTGATGAACTTGCAAGTCTTTGGGTT 458

301 CTGATGTGTCTTTGGGCAATGTTGCCAGAAAGGTATCTACCAAGATCTTTGGGAT 360

459 CTGATGTGTCTTTGGGCAATGTTGCCAGAAAGGTATCTACCAAGATCTTTGGGAT 518

361 GACAGGGTGTGCTCAAGGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420

519 GACAGGGTGTGCTCAAGGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 578

421 TTTGTCAGGCTGCTGCTCAACCAATCTCTCTGATGACCCCTGGGAGTTTCTTTGCACTGCA 480

579 TTTGTCAGGCTGCTGCTCAACCAATCTCTCTGATGACCCCTGGGAGTTTCTTTGCACTGCA 638

481 CTGCTCATTCGAGAGCTAAAGAGGTGTGCAACATGCGTGTGTTTAAAGTGGGCTCTTTC 540

639 CTGCTCATTCGAGAGCTAAAGAGGTGTGCAACATGCGTGTGTTTAAAGTGGGCTCTTTC 698

541 TCGGGCTCTGCTGGGCACTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600

699 TCGGGCTCTGCTGGGCACTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 758

601 CTGCTGTCTCATCTTCAACTTCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660

759 CTGCTGTCTCATCTTCAACTTCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 818

661 GCTGCTACCTGGGCTGTGATGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720

819 GCTGCTACCTGGGCTGTGATGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780

721 CAAGGCTCTGCTCAAGTCTTGGCCCAATGAGAAATGGGCTTCAATGGTGTGCCCTAT 938

879 CAAGGCTCTGCTCAAGTCTTGGCCCAATGAGAAATGGGCTTCAATGGTGTGCCCTAT 938

781 GTGTCCTCTGCTGTGGCAACAAGAAATCATCATGATCAAGATCAAGTGA 828

939 GTGTCCTCTGCTGTGGCAACAAGAAATCATCATGATCAAGATCAAGTGA 986

RESULT 2

AK123581

LOCUS 2852 bp mRNA linear PRI 09-SEP-2003

DEFINITION Homo sapiens cDNA FLJ14587 fis, clone CTONG2020638.

ACCESSION AK123581

VERSION AK123581.1 GI:34529163

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE

1 Ota, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Negahari, K., Masuno, Y., Nagai, K. and Isogai, T. NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 2852)

Isogai, T. and Yamamoto, J.

Direct Submission

Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kibarasu, Chiba 292-0818, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library

JOURNAL

REFERENCE

TITLE

AUTHORS

COMMENT

Db 939 GTGTCCTCTGTGTGCAACAGAAATCATCAGTCAAGATCACGTGA 986  
|||||

## RESULT 3

AC017081/c  
LOCUS AC017081 149462 bp DNA linear PRI 09-JAN-2002  
DEFINITION Homo sapiens BAC clone RP11-470J24 from 2, complete sequence.  
ACCESSION AC017081  
VERSION AC017081.8 GI:18093316  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM

## REFERENCE

1 (bases 1 to 149462)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Sulston, J.E. and Waterston, R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
99063792  
9847074

## REFERENCE

2 (bases 1 to 149462)  
Nguyen, C., Doebber, A. and Kozlowski, A.  
The sequence of Homo sapiens BAC clone RP11-470J24  
Unpublished (2001)  
3 (bases 1 to 149462)  
Waterston, R.H.

## REFERENCE

Direct Submission  
Submitted (09-DEC-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 149462)  
Waterston, R.

## REFERENCE

Direct Submission  
Submitted (09-JAN-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Jan 9, 2002 this sequence version replaced gi:14165368.  
----- Genome Center

## COMMENT

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
-----  
Center project name: H\_NH0470J24  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frenken, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute

(<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-310K15, 2000 bp overlap;  
The clone sequenced to the right is RP11-156A1, 2000 bp overlap.  
Actual start of this clone is at base position 190775 of  
RP11-310K15; actual end is at base position 28935 of RP11-156A1.

The sequence between 81601 and 81703 is covered only by a PCR product of clone DNA. Unresolved tandem repeat regions exist between 81548 and 83183, 126196 and 127426.

## FEATURES

| source        | Location/Qualifiers                                     |
|---------------|---|
| 1. .149462    | /organism="Homo sapiens"                                |
|               | /mol_type="genomic DNA"                                 |
|               | /db_xref="taxon:9606"                                   |
|               | /chromosome="2"   |
|               | /map="2"  |
|               | /clone="RP11-470J24"                                    |
|               | /clone_lib="RPCI-11"                                    |
|               | 141. .282   |
| repeat_region | /rpt_family="MIR"                                       |
| repeat_region | 581. .621   |
| repeat_region | /rpt_family="(TGGG)n"                                   |
| repeat_region | 965. .1037  |
| repeat_region | /rpt_family="GC-rich"                                   |
| misc_feature  | 1127. .1561   |
| misc_feature  | /note="match to EST AA056210 (NID:g1548612) zf62g06.s1" |
| repeat_region | 1771. .2127   |
| repeat_region | /note="match to EST AW614820 (NID:g7320006) hg80c08.x1" |
| repeat_region | 2055. .2195   |
| repeat_region | /rpt_family="MER1_type"                                 |
| repeat_region | 2216. .2419   |
| repeat_region | /rpt_family="MER1_type"                                 |
| repeat_region | 2604. .2844   |
| repeat_region | /rpt_family="MIR"                                       |
| misc_feature  | 2991. .3324   |
| misc_feature  | /note="match to EST R35058 (NID:g791959) yh86d09.r1"    |
| misc_feature  | 3207. .3758   |
| misc_feature  | /note="match to EST AW452004 (NID:g6992780) "           |
| repeat_region | 3486. .3527   |
| repeat_region | /rpt_family="AT-rich"                                   |
| repeat_region | 3501. .3745   |
| repeat_region | /note="match to EST AA054208 (NID:g1545131) zf51g03.r1" |
| repeat_region | 3981. .4017   |
| repeat_region | /rpt_family="T-rich"                                    |
| repeat_region | 4231. .4273   |
| repeat_region | /rpt_family="AT-rich"                                   |
| repeat_region | 5518. .5847   |
| repeat_region | /note="match to EST R80994 (NID:g857275) y194c08.r1"    |
| repeat_region | 5606. .5633   |
| repeat_region | /rpt_family="AT-rich"                                   |
| repeat_region | 5722. .5758   |
| repeat_region | /rpt_family="AT-rich"                                   |
| repeat_region | 5802. .5931   |
| repeat_region | /rpt_family="(TTTTA)n"                                  |
| repeat_region | 5808. .6116   |
| repeat_region | /rpt_family="Alu"                                       |
| repeat_region | 6153. .6244   |
| repeat_region | /rpt_family="MER1_type"                                 |
| repeat_region | 6965. .7000   |
| repeat_region | /rpt_family="AT-rich"                                   |
| repeat_region | 7020. .7094   |
| repeat_region | /rpt_family="MER1_type"                                 |
| repeat_region | 7982. .8006   |
| repeat_region | /rpt_family="AT-rich"                                   |
| repeat_region | 8000. .8436   |
| repeat_region | /rpt_family="L2"  |
| repeat_region | 8437. .8540   |
| repeat_region | /rpt_family="L1"  |
| repeat_region | 8541. .8696   |
| repeat_region | /rpt_family="L2"  |
| repeat_region | 8722. .9684   |



|   |     |   |     |
|---|-----|---|-----|
| Best Local Similarity 96.0%; Pred. No. 6.2e-196;                        |     | Matches 792; Conservative 0; Mismatches 30; Indels 3; Gaps 1; |     |
| QY  | 1   | ATGGCGCCCGCCACTGGTGGACACAGCTGCAGGCTGTAGTCGAGGTGACATGGTGC      | 60  |
| Db  | 1   | ATGGCGCCCGCCACTGGTGGACACAGCTGCAGGCTGGCAGCTCGAGGTGACATGGC      | 60  |
| QY  | 61  | GAGGCAACTACACATCGTGCCTGTATCGCGGAGTTCTAACACGATCAGCAATGTC       | 120 |
| Db  | 61  | GAGGCAACTACACATCGTGCCTGTATCGCGGAGTTCTATAACATGATCAGCAATGTC     | 120 |
| QY  | 121 | TTATTTTTCATTTTACCGCCCATCTGCATGCTCTGTTGTTGATGAGTATCAACATGCTTG  | 180 |
| Db  | 121 | TTATTTTTCATTTTACCGCCCATCTGCATGCTCTGTTGTTGATGAGTATCAACATGCTTG  | 180 |
| QY  | 181 | AACAGTGATCATCTAATCTGGAATCTTTTGGTTAGTGGGAATGGATCCGCTAC         | 240 |
| Db  | 181 | AACAGCGGCATCTAATCTGGAATCTTTTGGTTAGTGGGAATGGATCCGCTAC          | 237 |
| QY  | 241 | TTCCATTTTACCTTAGTTCTTGGGTGAGTCTTGTGATGAATCTTGAGTCTTTGGGT      | 300 |
| Db  | 238 | TTCCATTTTACCTTAGTTCTTGGGTGAGTCTTGTGATGAATCTTGAGTCTTTGGGT      | 297 |
| QY  | 301 | CTGATGCTGCTTTGCGCATGCTTCCCGAAGGTATCTACCAAGATCTTTGGGAT         | 360 |
| Db  | 298 | CTGATGCTGCTTTGCGCATGCTTCCCGAAGGTATCTACCAAGATCTTTGGGAT         | 357 |
| QY  | 361 | GACAGGGTAGTTCAGAGTGGTGTGATGCTCTGCTGCGGTACGACGTGCTGGCA         | 420 |
| Db  | 358 | GACAGGGTAGTTCAGAGTGGTGTGATGCTCTGCTGCGGTACGACGTGCTGGCA         | 417 |
| QY  | 421 | TTTGTCAAGCTGCGCATCAACAACTCTCTGTGATGACCTGGGAGTTCCTTGCACTGA     | 480 |
| Db  | 418 | TTTGTCAAGCTGCGCATCAACAACTCTCTGTGATGACCTGGGAGTTCCTTGCGTGA      | 477 |
| QY  | 481 | CTGCTCATCGCAGCTAAAGAGTGTGACAACTGCTGTTTAAAGCTGGGCTCTTC         | 540 |
| Db  | 478 | CTGCTCATCGCAGCTAAAGAGTGTGACAACTGCTGTTTAAAGCTGGGCTCTTC         | 537 |
| QY  | 541 | TGGGCTCTGCTGACCTGGCCCTGTTCTGCTGATCAGTGACCGAGCTTTCTGCGAG       | 600 |
| Db  | 538 | TGGGCTCTGCTGACCTGGCCCTGTTCTGCTGATCAGTGACCGAGCTTTCTGCGAG       | 597 |
| QY  | 601 | CTGCTGTATCTTCAACTTCCCTACCTGCTGATGCTGATGCGCATCTCATCTATGCTCT    | 660 |
| Db  | 598 | CTGCTGTATCTTCAACTTCCCTACCTGCTGATGCTGATGCGCATCTCATCTATGCTCT    | 657 |
| QY  | 661 | GCTGCTACCTGGGCTGTATGCTTTGCTACTTTGATGCTGCTCAGAGATCTCTGAG       | 720 |
| Db  | 658 | GCTGCTACCTGGGCTGTATGCTTTGCTACTTTGATGCTGCTCAGAGATCTCTGAG       | 717 |
| QY  | 721 | CAAGCCCTGTATCAAGTTCTGGCCCAATGAGAAATGGGCTTCAATGGTGTCCCTAT      | 780 |
| Db  | 718 | CAAGCCCTGTATCAAAATTTGGCCCAAGAGAAATGGGCTTCAATGGTGTCCCTAT       | 777 |
| QY  | 781 | GTGTCCTCTGTGTGCGCAACAGAAATCATCAGTCAAGATCAAG                   | 825 |
| Db  | 778 | GTGTCCTCTGTGTGCGCAACAGAAATCATCAGTCAAGATCAAG                   | 822 |
| RESULT 5  |     |   |     |
| AF282864  |     |   |     |
| LOCUS   |     |   |     |
| DEFINITION Mus musculus cancer related gene-liver 1 mRNA, complete cds. |     |   |     |
| ACCESSION AF282864  |     |   |     |
| VERSION AF282864.1  |     |   |     |
| KEYWORDS  |     |   |     |
| SOURCE Mus musculus (house mouse)                                       |     |   |     |
| ORGANISM  |     |   |     |
| REFERENCE   |     |   |     |
| AUTHORS   |     |   |     |
| Graveel, C.R., Jackoe, T., Madore, S.J., Holt, A.L. and Farnham, P.J.   |     |   |     |
| 1 (bases 1 to 4174)   |     |   |     |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;       |     |   |     |
| Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.      |     |   |     |
| TITLE   |     |   |     |
| Expression profiling and identification of novel genes in               |     |   |     |
| hepatocellular carcinomas   |     |   |     |
| Oncogene 20 (21), 2704-2712 (2001)                                      |     |   |     |
| JOURNAL MEDLINE   |     |   |     |
| 21313787  |     |   |     |
| PUBMED  |     |   |     |
| 11420682  |     |   |     |
| REFERENCE   |     |   |     |
| 2 (bases 1 to 4174)   |     |   |     |
| Graveel, C.R., Jackoe, T., Madore, S.J., Holt, A.L. and Farnham, P.J.   |     |   |     |
| Identification of genes deregulated in murine hepatocellular            |     |   |     |
| carcinomas using oligonucleotide microarrays and representational       |     |   |     |
| difference analysis   |     |   |     |
| Unpublished   |     |   |     |
| 3 (bases 1 to 4174)   |     |   |     |
| Farnham, P.J. and Graveel, C.R.   |     |   |     |
| Direct Submission   |     |   |     |
| Submitted (27-JUN-2000) Oncology, University of Wisconsin, 1400         |     |   |     |
| University Avenue, Madison, WI 53706, USA                               |     |   |     |
| JOURNAL   |     |   |     |
| FEATURES  |     |   |     |
| source  |     |   |     |
| 1..4174   |     |   |     |
| Location/Qualifiers   |     |   |     |
| /organism="Mus musculus"  |     |   |     |
| /mol_type="mRNA"  |     |   |     |
| /db_xref="taxon:10090"  |     |   |     |
| /tissue_type="hepatocellular carcinoma"                                 |     |   |     |
| 35..862   |     |   |     |
| CDS   |     |   |     |
| /note="CRG-L1; putative"  |     |   |     |
| /codon_start=1  |     |   |     |
| /product="cancer related gene-liver 1"                                  |     |   |     |
| /protein_id="AAL40408.1"  |     |   |     |
| /db_xref="GI:17529684"  |     |   |     |
| Translation="MGAPHWMDHLRAGSEVDWCDNVTIVPAIEFYNTISNVLFFI                  |     |   |     |
| LPPICMLFRQYATCFNSGIYLIWLLVVGISVYFPHATISFLQGLDELAILVLM                   |     |   |     |
| CALAMFPRYLPKIFRNDGRFRKAVCVLSAITTCLAFIPAINNINISLGLPCTA                   |     |   |     |
| LLVLAELKRCNDNRVFKLGFSLGMLWTLALFCWISQACELLSFHFYLVHCVHILI                 |     |   |     |
| CLASYLGVCVFAYFDAASEIPEQGVIRFEPSEKWAFTIGVPYVLSLLCAHKKSVPKIT"             |     |   |     |
| ORIGIN  |     |   |     |
| Query Match 80.3%; Score 664.8; DB 10; Length 4174;                     |     |   |     |
| Best Local Similarity 87.7%; Pred. No. 1.1e-169;                        |     |   |     |
| Matches 726; Conservative 0; Mismatches 102; Indels 0; Gaps 0;          |     |   |     |
| QY  | 1   | ATGGCGCCCGCCACTGGTGGACACAGCTGCAGGCTGGTGTAGTCGAGGTGACATGGTGC   | 60  |
| Db  | 35  | ATGGCGCCCGCCACTGGTGGACACAGCTGCAGGCTGGCAGTTCGAGGTGATTTGGTGC    | 94  |
| QY  | 61  | GAGGCAACTACACATCGTGCCTGCTATGCGCGAGTTCTACAAACGATCAGCAATGTC     | 120 |
| Db  | 95  | GAGGCAACTACACATCGTGCCTGCCATTCGCGAGTTCTACAAACGATCAGCAACGTC     | 154 |
| QY  | 121 | TTATTTTTCATTTTACCGCCCATCTGCATGCTCTGTTTGTGATGAGTATCAACATGCTTG  | 180 |
| Db  | 155 | TTGTTTTTCATTTTACCTCCCATCTGCATGCTGTTTCCGCGAGTACGCAACGCTGCTTC   | 214 |
| QY  | 181 | AACAGTGATCATCTAATCTGGAATCTTTTGGTTAGTGGGAATGGATCCGCTAC         | 240 |
| Db  | 215 | AACAGCGCATCTAATCTAATGAGCGCTCTAGTTGTAGTGGGATTTGGATCTGCTAC      | 274 |
| QY  | 241 | TTCCATTTTACCTTAGTTCTTGGGTGAGTCTTGTATGAACTTGCAAGTCTTTGGGTT     | 300 |
| Db  | 275 | TTCCATGCAACGCTGAGTTTCTGGGTGAGTCTTGTATGAACTTGCAAGTCTTTGGGTT    | 334 |
| QY  | 301 | CTGATGCTGCTTTGGCCATGTTTCCCGAAGGTATCTACCAAGATCTTTTCGGAT        | 360 |
| Db  | 335 | CTGATGCTGCTTTGGCCATGTTTCCCGAAGGTATTTACCAAGATCTTTTCGGAT        | 394 |
| QY  | 361 | GACAGGGTAGTTCAGAGTGGTGTGCTGCTGCTGCTGCGGTGTACGAGCTGCTGGCA      | 420 |
| Db  | 395 | GACAGGGTAGTTCAGAGTGGTGTGCTGCTGCTGCTGCGGTGTACGAGCTGCTGGCA      | 454 |
| QY  | 421 | TTTGTCAAGCTGCGCATCAACAACTCTCTGTATGACCTGGGAGTTCCTTGCACTGCA     | 480 |
| Db  | 455 | TTTGTCAAGCTGCGCATCAACAACTCTCTGTATGACCTGGGAGTTCCTTGCACTGCG     | 514 |
| QY  | 481 | CTGCTCATCGCAGCTAAAGAGTGTGACAACTGCTGTTTAAAGCTGGGCTCTTC         | 540 |

|     |   |     |
|-----|---|-----|
| 515 | CTGCTTTGCCAGCTGAAGGTGTGACAAATGTGCTGTGTTTAAGCTGGGCTCTTC        | 574 |
| 541 | TCGGGCTCTGGTGGACCCCTGGCCCTGTTCTGTCTGGATCAGTGACCGAGCTTTCTGGGAG | 600 |
| 575 | TCTGGCTCTCTGGTGGACTCTGGCTCTCTTCTGTCTGGATCAGGACCAAGCCTTCTGTGAG | 634 |
| 601 | CTGCTGTATCCCTTCAACTTCCCCTACTGTCACATGCATGTGGGCATCCCTCATCTGCCTT | 660 |
| 635 | CTGCTCTCTCTCTTCACTTCCCCTACTGTCACATGTGTGGCATAATTCTCATCTGCCTT   | 694 |
| 661 | GCTGCCCTACTGGGCTGTGATGTTTGCCTACTTTGATGCTGCTCAGAGATTCCTGAG     | 720 |
| 695 | GCTTCGTACCTCGGCTGTGTGTGCTTGCCCTACTTTGATGCTGCTCAGAGATACCTGAG   | 754 |
| 721 | CAAGGCCCTGTCATCAAGTTCTGCGCCCAATGAGAAATGGGCGCTTCATTGGTGTCCCTAT | 780 |
| 755 | CAAGGTCAGTCATCAGATTCTGCGCCAGCGAGAAATGGGCTTTATTGGTGTCCCTTAT    | 814 |
| 781 | GTCTGCCCTCTGTGTGCCAACAGAAATCATCAGTCAAGATCACTGTA               | 828 |
| 815 | GTCTGCCCTTCTGTGTGCCCAAGAAAGTCGCGAGTCAAGATCAGGTGA              | 862 |

|            |   |         |  |
|------------|---|---------|--|
| RESULT 6   |   |         |  |
| BC059819   |   |         |  |
| LOCUS      | BC059819                                  | 4078 bp | mRNA linear ROD 21-OCT-2003                              |
| DEFINITION | Mus musculus cancer related gene-liver 1, |         | mRNA (cdna clone MGC:69583 IMAGE:6839525), complete cds. |



Db 436 GAC----- 438  
QY 421 TTTGTCAAGCCTGCATCAACAAACATCTCTCTGATGACCCCTGGAGTTCTTTGCACTGCA 480  
Db 439 ----- 438  
QY 481 CTGCTCATCGCAGACTAAAGAGGTGTGACAAACATGCGTGTGTTTAAAGTGGCCTCTTC 540  
Db 439 -----AGGTGTGACATGTGCGTGTGTTTAAAGTGGCCTCTTC 477  
QY 541 TCGGCGCTCTGTGACCTGCGCCTGTCTCTGATCAGTACCGAGCTTTCTGCGAG 600  
Db 478 TCTGGCCTCTGGTGAATCTGCTCTCTCTGATCAGGACCAAGCCTCTCTGTGAG 537  
QY 601 CTGCTGTCATCTTCAACTTCCCTTACCTGCACTGATGTCGACATCTCTATCTGCTTC 660  
Db 538 CTGCTCTCTCTTCACTTCCCTTACCTGCACTGATGTCGACATTTCTCATCTGCTTC 597  
QY 661 GCTGCTACCTGGGTGTGATGCTTTGCTTACCTTGTGATGCTGCTCAGAGATTCCTGAG 720  
Db 598 GCTTGTGATACCTGGGTGTGATGCTTGTGCTTGTGCTTGTGATGCTGCTCAGAGATTCCTGAG 657  
QY 721 CAAGCGCTGTGATCAAGTCTTGGCCCAATGAGAAATGAGAAATGAGTGTGCTTCTTAT 780  
Db 658 CAAGTTCAGTCAATCAGATTTGCGCCAGGAGAAATGGGCTTTTATTGGTGTCCCTTAT 717  
QY 781 GTGTCCCTCTGTGTGCAACAAGAAATCATCAGTCAAGATCACGTGA 828  
Db 718 GTGTCCCTCTGTGTGCCCCAAGAAGTGGCAGTCAAGATCACGTGA 765

RESULT 7  
LOCUS CQ744096 315 bp DNA linear PAT 03-FEB-2004  
DEFINITION Sequence 30030 from Patent WO02068579.  
ACCESSION CQ744096  
VERSION CQ744096.1 GI:42358801

KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.  
TITLE Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof  
JOURNAL Patent: WO 02068579-A 30030 06-SEP-2002;  
PE Corporation (NY) (US)

FEATURES  
source Location/Qualifiers  
1..315  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 38.0%; Score 315; DB 6; Length 315;  
Best Local Similarity 100.0%; Pred. No. 1.5e-74;  
Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 ATGGTGTGTTTAAAGCTGGGCTCTTCTCGGCGCTCTGTTGAGACCTGGCCTGTGTCG 573  
Db 1 ATGGTGTGTTTAAAGCTGGGCTCTTCTCGGCGCTCTGTTGAGACCTGGCCTGTGTCG 60  
QY 574 TGGATCAGTACCGAGCTTTCTGGAGCTGTGTCATCTTCAACTCCCTACCTGCAC 633  
Db 61 TGGATCAGTACCGAGCTTTCTGGAGCTGTGTCATCTTCAACTCCCTACCTGCAC 120  
QY 634 TGCATGTGGCACATCTCTCATCTGCTTGTGCTACCTGGGCTGTGATGCTTTGCCTAC 693  
Db 121 TGCATGTGGCACATCTCTCATCTGCTTGTGCTACCTGGGCTGTGATGCTTTGCCTAC 180  
QY 694 TTTGATGTGCTCAGAGATTCCTGAGCAAGCCCTGTGTCATCAAGTTCTGGCCCAATGAG 753  
Db 181 TTTGATGTGCTCAGAGATTCCTGAGCAAGCCCTGTGTCATCAAAATTTGGCCCAATGAG 240  
QY 754 AAATGGGCTTCAATGGTGTCCCTATGTGTCCTCTGTCGCAACAAGAAATCATCA 813  
Db 241 AAATGGGCTTCAATGGTGTCCCTATGTGTCCTCTGTCGCAACAAGAAATCATCA 300

QY 694 TTTGATGTGCTCAGAGATTCCTGAGCAAGCCCTGTGTCATCAAGTTCTGGCCCAATGAG 753  
Db 181 TTTGATGTGCTCAGAGATTCCTGAGCAAGCCCTGTGTCATCAAGTTCTGGCCCAATGAG 240  
QY 754 AAATGGGCTTCAATGGTGTCCCTATGTGTCCTCTGTCGCAACAAGAAATCATCA 813  
Db 241 AAATGGGCTTCAATGGTGTCCCTATGTGTCCTCTGTCGCAACAAGAAATCATCA 300  
QY 814 GTCAAGATCACGTGA 828  
Db 301 GTCAAGATCACGTGA 315

RESULT 8  
LOCUS CQ736905 315 bp DNA linear PAT 03-FEB-2004  
DEFINITION Sequence 22839 from Patent WO02068579.  
ACCESSION CQ736905  
VERSION CQ736905.1 GI:42333763

KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.  
TITLE Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof  
JOURNAL Patent: WO 02068579-A 22839 06-SEP-2002;  
PE Corporation (NY) (US)

FEATURES  
source Location/Qualifiers  
1..315  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 37.3%; Score 308.6; DB 6; Length 315;  
Best Local Similarity 98.7%; Pred. No. 8.4e-73;  
Matches 311; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 514 ATGGTGTGTTTAAAGCTGGGCTCTTCTCGGCGCTCTGTTGAGACCTGGCCTGTGTCG 573  
Db 1 ATGGTGTGTTTAAAGCTGGGCTCTTCTCGGCGCTCTGTTGAGACCTGGCCTGTGTCG 60

QY 574 TGGATCAGTACCGAGCTTTCTGGAGCTGTGTCATCTTCAACTCCCTACCTGCAC 633  
Db 61 TGGATCAGTACCGAGCTTTCTGGAGCTGTGTCATCTTCAACTCCCTACCTGCAC 120

QY 634 TGCATGTGGCACATCTCTCATCTGCTTGTGCTACCTGGGCTGTGATGCTTTGCCTAC 693  
Db 121 TGCATGTGGCACATCTCTCATCTGCTTGTGCTACCTGGGCTGTGATGCTTTGCCTAC 180

QY 694 TTTGATGTGCTCAGAGATTCCTGAGCAAGCCCTGTGTCATCAAGTTCTGGCCCAATGAG 753  
Db 181 TTTGATGTGCTCAGAGATTCCTGAGCAAGCCCTGTGTCATCAAAATTTGGCCCAATGAG 240

QY 754 AAATGGGCTTCAATGGTGTCCCTATGTGTCCTCTGTCGCAACAAGAAATCATCA 813  
Db 241 AAATGGGCTTCAATGGTGTCCCTATGTGTCCTCTGTCGCAACAAGAAATCATCA 300

QY 814 GTCAAGATCACGTGA 828  
Db 301 GTCAAGATCACGTGA 315

RESULT 9  
LOCUS CQ736906 315 bp DNA linear PAT 03-FEB-2004  
DEFINITION Sequence 22840 from Patent WO02068579.  
ACCESSION CQ736906  
VERSION CQ736906.1 GI:42333767

```
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Venter, C.J., Adams M.C., Li, P.W. and Myers, E.W.
Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 22840 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
    source
        1..315
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
ORIGIN
    Query Match 37.3%; Score 308.6; DB 6; Length 315;
    Best Local Similarity 98.7%; Pred. No. 8.4e-73;
    Matches 311; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 514 ATGCGTGTGTTAAGCTGGGCTCTTCTCGGGCTCTGTGGACCTGTGGCCCTGTTCTGC 573
Db 1 ATGCGTGTGTTAAGCTGGGCTCTTCTCGGGCTCTGTGGACCTGTGGCCCTGTTCTGC 60
QY 574 TGGATCAGTACCGAGCTTTCTGGAGCTGTGTGTCATCTTCAACTTCCCGTACCTGCAC 633
Db 61 TGGATCAGTACCGAGCTTTCTGGAGCTGTGTGTCATCTTCAACTTCCCGTACCTGCAC 120
QY 634 TGCATGGGCACATCCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 693
Db 121 TGCATGGGCACATCCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
QY 694 TTTGATGCTGCTCAGAGATCTCTGAGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 753
Db 181 TTTGATGCTGCTCAGAGATCTCTGAGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
QY 754 AATGGGCTTCATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 813
Db 241 AATGGGCTTCATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
QY 814 GTCAAGATCAGTGA 828
Db 301 GTCAAGATCAGTGA 315
RESULT 10
AF323976 1163 bp mRNA linear INV 01-JAN-2002
LOCUS Drosophila melanogaster brain washing (bwa) mRNA, complete cds.
DEFINITION AF323976
ACCESSION AF323976.1 GI:18028134
VERSION AF323976.1
KEYWORDS
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
AUTHORS Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
TITLE Ephydroidea; Drosophilidae; Drosophila.
JOURNAL Pascual, A., Boquet, I. and Preat, T.
Submitted (24-NOV-2000) Institut de Neurobiologie Alfred Fessard,
Avenue de la Terrasse, Bat 32/33, Gif-sur-Yvette 91190, France
FEATURES
    source
        1..1163
            /organism="Drosophila melanogaster"
            /mol_type="mRNA"
            /db_xref="taxon:7227"
            /chromosome="2"
            /map="38A"
            1..1163
gene
```

## CDS

```
/gene="bwa"
195..1046
/genes="bwa"
/codon_start=1
/product="brain washing"
/protein_id="AAL55991.1"
/db_xref="GI:18028135"
/translation="MGMMGGGLDIYMAWEQRLPGSSPDWCEGNLYLISSNIAEFV
NTPSNFLFILLPVLIMLFKEYGRFVPGHIVLWLLIVVGLSSMYHATLSLIGQLL
DEAILWVMAAFSLFPRKYFKVNDKRTESWMLLSAIAATGLSWKPIVNAFV
LMFMSVPTVMVLYTELQVSDQRYRLSTVMAVAVFCVNDIRLFCFCAWSINFP
YLHGFWHIFIFIAAYTVLVLFAYFYVESELPPQRLKLYWPKNEFEFGIPFISIRNP
KALNTI"
ORIGIN
    Query Match 29.0%; Score 239.8; DB 3; Length 1163;
    Best Local Similarity 57.6%; Pred. No. 5.1e-54;
    Matches 430; Conservative 0; Mismatches 317; Indels 0; Gaps 0;
QY 19 TGGGACCACTGTCAGGCTGTAGCTCGAGTGGACTGGTGGAGGACAACTACACCATC 78
Db 243 TGGGAGCACTTAAGACCCGGAAGCTGCCGCTGACTGGTGGAGGACAACTACTTGATT 302
QY 79 GTGCCTGCTATCGCCGAGTTCTACACACGATCAGCAATGCTTATTTTCAATTTTACCG 138
Db 303 TCATCCAACTGCGCGAGTTCTGTAACACGTTTAGCAACTTCTGTTTCACTTCTAC 362
QY 139 CCATCTGCAATGCTGTTGTTGATGAGTATGCAACATGCTTGAACAGTACATCTACTTA 198
Db 363 CCGCTCTGATAATGCTCTTCAAGGAGTACGAGCGCTTGTGACGCCCGGAAATCCAGTC 422
QY 199 ATCTGGACTCTTTTGGTTGTTAGTGGGAATTTGGATCCGCTACTTCCATTTTACCTTAGT 258
Db 423 ATCTGGGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 482
QY 259 TTCTTGGGTGATGCTTGTATGAACTTGCAGTCTCTTGGGCTTCTGATGTGCTGCTTGGCC 318
Db 483 CTGATTTGGCAGCTGCTGGAGCAACTGGCCATCTCTGGGCTCTTCACTGGCGGCTTTTCG 542
QY 319 ATGTGTTTCCCAGAGGATATCTACCAAGATCTTTCGGAATGACAGGGGTAGGTTCAAG 378
Db 543 CTCTTCTATCCGAAGCGATATCTCCAAAGTTCGTGAAAAACGATCGCAAAACCTTCAGT 602
QY 379 GTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 438
Db 603 TGGCTCATGCTCTTGTGCGCGATTTGTCGACGGGCTTGTGCTGCTGCTGCTGCTGCTGCT 662
QY 439 AACAAATCTCTCTGATGACCTCGGAGTTCTTTCGACTGCTGCTGCTGCTGCTGCTGCTGCT 498
Db 663 AACGCTTTGTTCTCATGTTTCATGAGTGTGCCGACCATGTTGTTGCTTACACAGAGCTG 722
QY 499 AAGAGTGTGACAACTGCTGTTTAAAGCTGGGCTCTTCTTCTGGGCTCTTCTGCTGAGCC 558
Db 723 CAGAGAGTTAGTAGCACCAGAGGGTCTACCGCTCGGCATCCGATCGACGCGCTGCGGCA 782
QY 559 CTGSCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 618
Db 783 GTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 842
QY 619 TTCCCTTACCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 678
Db 843 TTTCCATACCTGCAAGGCTTCTGCGACATATTTATTTTATAGCGGCTTACACGGTCTG 902
QY 679 GTATGCTTTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 738
Db 903 GTGCTCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 962
QY 739 TTCTGGGCCCAATGAGAAATGGGCTTC 765
Db 963 TACTGGCCGAGAACGAGTTCGAGTTC 989
```

```

RESULT 11
AY071232      2318 bp mRNA linear INV 20-DEC-2001
Drosophila melanogaster RE26924 full length cDNA.
ACCESSION
AY071232
AY071232.1 GI:17945605
FLI CDNA.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 2318)
Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J.,
Champe, R., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E.,
George, R., Gonzalez, M., Guarin, H., Kronmiller, B., Li, P., Liao, G.,
Miranda, A., Mungall, C. J., Nunco, J., Pacleab, J., Paragas, V., Park, S.,
Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Rubin, G. M.
and Celniker, S.
Direct Submission
Submitted (18-DEC-2001) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.
FEATURES
source
1..2318
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/map="38B4-38B5"
1..2318
/gene="CG13969"
/note="alignment with genomic scaffold AB003665. gene does
not completely overlap longest ORF"
/db_xref="FLYBASE:FBgn0032842"
404..1255
/gene="CG13969"
/note="Longest ORF"
/codon_start=1
/product="RE26924p"
/protein_id="AAL48854.1"
/db_xref="GI:17945606"
/db_xref="FLYBASE:FBgn0032842"
/translation="MGNGGGGLLDIYAMANEHLRPGSPVDWCEGNVLSSNTAEFV
NFSNFLLIPPLVIMLKEYGRFVTPGIHVIWLVVLGLSSMYFHTLSLIGQLL
DELAIVFMFAVLSFYPRYKPKNDKRTFSWMLLSAATGSLWPKPIVNAFV
LMFMSVPTVMVTELQDSVQVRVLRIGSTTVMAVAVFCWINDRIFCEAWSSINFP
YLHGFWHIFIFIAATVVLVLFAYFVSELSRQLPPLKYWKNEFEFGIPISIRNPG
KALNTI"
ORIGIN
Query Match 28.8%; Score 238.2; DB 3; Length 2318;
Best Local Similarity 57.4%; Pred. No. 1.5e-53;
Matches 429; Conservative 0; Mismatches 318; Indels 0; Gaps 0;

```





|    | Matches | 412;  | Conservative | 0; | Mismatches | 359; | Indels | 3; | Gaps | 1; |
|----|---------|---|--------------|----|------------|------|--------|----|------|----|
| QY | 40      | AGCTCGGAGTGACTGCTGCGAGGACAACTACACCATCGTGCCTGCTATCGCGAGTTC     | 99           |    |            |      |        |    |      |    |
| Db | 25      | AGCTCGGAGTGACTGCTGCGAGGACAACTTCAGTACTCGGAGCTGCTGCGAGTTC       | 84           |    |            |      |        |    |      |    |
| QY | 100     | TACAACACGATACGCAATGCTTATTTTTCATTTTACCGCCCATCTGCATGCTGCTGTTT   | 159          |    |            |      |        |    |      |    |
| Db | 85      | TACAACACGATACGCAATGCTTATTTTTCATTTTACCGCCCATCTGCATGCTGCTGTTT   | 144          |    |            |      |        |    |      |    |
| QY | 160     | GATGAGTATGCAACATGCTTTGAAACAGTACATCTACTTAATCTGGAATCTTTTGGTTGTA | 219          |    |            |      |        |    |      |    |
| Db | 145     | CACCCGTATGCCAGAGCGCTCCGCTACATTTAGCTTCTGCTGCTCTTCTCATGATC      | 204          |    |            |      |        |    |      |    |
| QY | 220     | GTGGGAATGGATCCGCTACTTCCATTTTACCTTAGTTTCTTGGGTAGATGCTTGAT      | 279          |    |            |      |        |    |      |    |
| Db | 205     | ATAGGCTGTCTTCCATGATTTTCCATGACGCTCAGCTTCTTGGGCCAGCTGCTGGAC     | 264          |    |            |      |        |    |      |    |
| QY | 280     | GAACCTTGCACTCCCTTTGGGTTCTGATGTGTGCTTTTGGCCATGTGTTCCCGAGAGGTAT | 339          |    |            |      |        |    |      |    |
| Db | 265     | GAGATGCCATCTGTGCTCTTGGGAGTGGCTATAGCATATGATGCTCCCGCTGCTAT      | 324          |    |            |      |        |    |      |    |
| QY | 340     | CTACCAAGATCTTTTCGGAATGACAGGGTAGGTTCAAGTGTGTGCTGCTGCTCTCT      | 399          |    |            |      |        |    |      |    |
| Db | 325     | TTCCCTCTCTTCTTGGGGGAAACAGGTCCAGTTTCATCGCCTGCTTTCATCACCAC      | 384          |    |            |      |        |    |      |    |
| QY | 400     | GGGTTACGAGTGCCTGGCATTTTGTCAAGCTGCCATCAACAACATCTCTCTGATGACC    | 459          |    |            |      |        |    |      |    |
| Db | 385     | GTGGTCAGCACCTTCTGTCTTCTGCTGCGGCCCAACGCTCAACGCTTACGCCCTCAAC    | 444          |    |            |      |        |    |      |    |
| QY | 460     | CTGGGAGTTCCTTGCACTGCACCTCATCGCAGAGCTAAAGAGGTGTGACACATGCGT     | 519          |    |            |      |        |    |      |    |
| Db | 445     | ATTGCCCTGCACATCTCTACATCGTGTGCCAGAGTACAGAGAGACCAGCAATAGGAG     | 504          |    |            |      |        |    |      |    |
| QY | 520     | GTGTTAAGCTGGGCTCTTCTCGGGCTCTGTGGGACCTGTGGACCTTCTGCTGATC       | 579          |    |            |      |        |    |      |    |
| Db | 505     | CTTCGGCACCTGATTTGAGGTCTCGTGGTTTATGGGCTGTGCTCTGACCGAGCTGGATC   | 564          |    |            |      |        |    |      |    |
| QY | 580     | AGTACCGAGCTTCTCGGAGCTGCTGTATCTTCACTTCCCTACCTGCACTGCATG        | 639          |    |            |      |        |    |      |    |
| Db | 565     | AGTACCGCTGTGCTTTGCACTTCTGCAAGGATTCATTTCTTCTGACAGCATC          | 624          |    |            |      |        |    |      |    |
| QY | 640     | TGGCACATCTCTATCTGCTGCTTGTGCTTACCTGGGCTGTGATGCTTTGCTTCTTGTAT   | 699          |    |            |      |        |    |      |    |
| Db | 625     | TGGCATGTGTCTATCAGCATCACCTTCCCTTATGGCATGGTCCACATGGCTTGGTGGAT   | 684          |    |            |      |        |    |      |    |
| QY | 700     | GCTGCTTCAAGATTTCTGAGCAAGGCTGTGTATCAAGTTCTGGCCCAATGAGAAATGG    | 759          |    |            |      |        |    |      |    |
| Db | 685     | GCCAACTATGAGATGCCAGGTGAAACCTCAAGTCCGCTACTGGCTCGGGACAGTTGG     | 744          |    |            |      |        |    |      |    |
| QY | 760     | GCCTTCATTGGTCCCTTATGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG    | 813          |    |            |      |        |    |      |    |
| Db | 745     | CC---CGTGGGCTGCTTACGTTGAAATCGGGGTGATGACAAGGACTGCTGA           | 795          |    |            |      |        |    |      |    |

Search completed: November 21, 2004, 03:07:27  
Job time : 3363.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 20:48:16 ; Search time 351.193 Seconds  
(without alignments)  
12376.463 Million cell updates/sec

Title: US-10-017-410-3

Perfect score: 828

Sequence: 1 atgggcgcgcgcactggtg.....catcagtcagatcacgtga 828

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 23Sep04.\*

1: Geneseqn1980s.\*

2: Geneseqn1990s.\*

3: Geneseqn2000s.\*

4: Geneseqn2001as.\*

5: Geneseqn2001bs.\*

6: Geneseqn2002as.\*

7: Geneseqn2002bs.\*

8: Geneseqn2003as.\*

9: Geneseqn2003bs.\*

10: Geneseqn2003cs.\*

11: Geneseqn2003ds.\*

12: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 828   | 100.0       | 828    | 6     | ABS54149           |
| 2          | 813.6 | 98.3        | 4202   | 6     | ACC50961 Human cDN |
| 3          | 813.6 | 98.3        | 4212   | 11    | ADN38833 Cancer/an |
| 4          | 761   | 91.9        | 822    | 4     | AH48641 Human cer  |
| 5          | 664.8 | 80.3        | 4175   | 6     | ABS54148           |
| 6          | 486.2 | 58.7        | 1623   | 10    | ADI21948           |
| 7          | 414.8 | 50.1        | 1527   | 10    | ADL06640 Human 3T3 |
| 8          | 414.8 | 50.1        | 1527   | 10    | ADL06642 Human 3T3 |
| 9          | 313   | 37.8        | 487    | 10    | ADI21468           |
| 10         | 185   | 22.3        | 1215   | 4     | AKS1566            |
| 11         | 185   | 22.3        | 1631   | 4     | AAK52550           |
| 12         | 183.4 | 22.1        | 792    | 4     | AAH48639           |
| 13         | 147.8 | 17.9        | 537    | 4     | ABL24043           |
| 14         | 95.6  | 11.5        | 4071   | 4     | ABL30432           |
| 15         | 75    | 9.1         | 4758   | 4     | ABL24042           |
| 16         | 52.2  | 6.3         | 35425  | 10    | ADC87616           |
| 17         | 44.8  | 5.4         | 2000   | 8     | ADA71938           |
| 18         | 41.2  | 5.0         | 2000   | 8     | ADA71938           |
| 19         | 40.2  | 4.9         | 354    | 10    | ABX74425           |
| 20         | 40.2  | 4.9         | 636    | 6     | ABL90053 Human cDN |
| 21         | 40.2  | 4.9         | 748    | 4     | AAH07955 Human cDN |

|    |      |     |       |    |           |
|----|------|-----|-------|----|-----------|
| 22 | 40.2 | 4.9 | 801   | 4  | AAH48640  |
| 23 | 40.2 | 4.9 | 1063  | 6  | ABZ11993  |
| 24 | 40.2 | 4.9 | 1063  | 12 | ADM44511  |
| 25 | 40.2 | 4.9 | 2058  | 4  | AAH15072  |
| 26 | 40.2 | 4.9 | 2744  | 3  | AAA96499  |
| 27 | 40.2 | 4.9 | 4646  | 5  | ABV27900  |
| 28 | 40.2 | 4.9 | 4646  | 5  | ABV28037  |
| 29 | 40.2 | 4.9 | 4646  | 5  | ABV22202  |
| 30 | 40.2 | 4.9 | 4646  | 5  | ABV22064  |
| 31 | 40.2 | 4.9 | 6114  | 4  | ABL24565  |
| 32 | 40.2 | 4.9 | 9287  | 4  | ABL24564  |
| 33 | 38.8 | 4.7 | 332   | 12 | ACH93876  |
| 34 | 38.8 | 4.7 | 579   | 12 | ACH80176  |
| 35 | 38.8 | 4.7 | 1057  | 8  | ACA57321  |
| 36 | 38.8 | 4.7 | 2271  | 4  | AAH14050  |
| 37 | 38.8 | 4.7 | 2591  | 6  | ABA93743  |
| 38 | 38.8 | 4.7 | 3383  | 6  | ABQ60889  |
| 39 | 38.8 | 4.7 | 34989 | 10 | ADC86916  |
| 40 | 38.6 | 4.7 | 639   | 5  | AA579600  |
| 41 | 38.6 | 4.7 | 10732 | 3  | AAA10594  |
| 42 | 38   | 4.6 | 2800  | 4  | AAAD07131 |
| 43 | 38   | 4.6 | 2803  | 4  | AAAD07132 |
| 44 | 38   | 4.6 | 2805  | 4  | AAAD07130 |
| 45 | 37.4 | 4.5 | 1551  | 3  | AAZ98176  |

## ALIGNMENTS

RESULT 1  
ABS54149  
ID ABS54149 standard; cDNA; 828 BP.  
XX  
AC ABS54149;  
XX  
DT 25-NOV-2002 (first entry)  
XX  
DE Human cDNA encoding a liver tumour marker protein, CRG-L1.  
XX  
KW Human; ss; gene; liver cancer; liver tumour; CRG-L1;  
KW hepatocellular cancer; chromosome 9p.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..828  
FT /\*tag= a  
FT /product= "Liver tumour marker protein, CRG-L1"  
XX  
PN US2002115094-A1.  
XX  
PD 22-AUG-2002.  
XX  
PF 14-DEC-2001; 2001US-00017410.  
XX  
PR 14-DEC-2000; 2000US-0255674P.  
XX  
PA (FARN/) FARNHAM P J.  
PA (GRAV/) GRAVEEL C R.  
XX  
PI Farnham PJ, Graveel CR;  
XX  
DR WPI; 2002-706409/76.  
DR P-PSDB; ARG32881.  
XX  
PT Novel polypeptide designated as CRG-L1, useful as diagnostic marker for  
PT liver cancer, is differentially expressed in liver tumors relative to  
PT normal liver tissues.  
XX  
CC Claim 2; Page 8; 11pp; English.  
XX  
CC The invention relates to a polypeptide designated as CRG-L1, which is  
XX differentially expressed in liver tumours relative to normal expression





|    |     |  |     |
|----|-----|--|-----|
| Qy | 181 | AACAGTGCACATCTACTTAATCTGGACTCTTTTGGTTGTAGTGGGAATTCGGATCCGCTCTAC  | 240 |
| Db | 257 | AACAGTGGCATCTACTTAATCTGGACTCTTTTGGTTGTAGTGGGAATTCGGATCCGCTCTAC   | 316 |
| Qy | 241 | TTCCATTTTACCCCTTAGTCTTTCTTGGGTGAGATGCTTGATGAATTCGACGTCTTTGGGGTT  | 300 |
| Db | 317 | TTCCATGCAACCCCTTAGTCTTTCTTGGGTGAGATGCTTGATGAATTCGACGTCTTTGGGGTT  | 376 |
| Qy | 301 | CTGATGTGTGCTTTTGGCCCATGTGGTTTCCCGAAGGTACTTACCAAGATCTTTTCGGAAAT   | 360 |
| Db | 377 | CTGATGTGTGCTTTTGGCCCATGTGGTTTCCCGAAGGTACTTACCAAGATCTTTTCGGAAAT   | 436 |
| Qy | 361 | GACAGGGGTAGTGTCAAGTGGTGTGTCAGTGCCTGTCTGCGGTTTACGACGTGCCTGGCA     | 420 |
| Db | 437 | GACCGGGTAGTGTCAAGTGGTGTGTCAGTGCCTGTCTGCGGTTTACGACGTGCCTGGCA      | 496 |
| Qy | 421 | TTTCTCAAGCCTGCGCATCAACAAACATCTCTCTGATGACCCCTGGGAGTTTCTTTGCACTGCA | 480 |
| Db | 497 | TTTGTCAAGCCTGCGCATCAACAAACATCTCTCTGATGACCCCTGGGAGTTTCTTTGCACTGCA | 556 |
| Qy | 481 | CTGCTCATCGCAGAGCTTAAGAGGTGTGACAAACATGCGTGTTTAAAGCTGGGCGCTCTTC    | 540 |
| Db | 557 | CTGCTCATCGCAGAGCTTAAGAGGTGTGACAAACATGCGTGTTTAAAGCTGGGCGCTCTTC    | 616 |
| Qy | 541 | TCGGGCGCTCTGGTGGACCGCTGGCCCTGTTCTCTGATCGACCGAGCTTTCTTGCAG        | 600 |
| Db | 617 | TCGGGCGCTCTGGTGGACCGCTGGCCCTGTTCTCTGATCGACCGAGCTTTCTTGCAG        | 676 |
| Qy | 601 | CTGCTGTCACTCTTCAACTTCCCTACTCGCACTGCATGTGGCACATCCTCATCTGCCTT      | 660 |
| Db | 677 | CTGCTGTCACTCTTCAACTTCCCTACTCGCACTGCATGTGGCACATCCTCATCTGCCTT      | 736 |
| Qy | 661 | GCTGCCCTACCTGGGCTGTGTATGTTTGGCTACTTTGATGCTGCGCTCAGAGATTCTGTAG    | 720 |
| Db | 737 | GCTGCCCTACCTGGGCTGTGTATGTTTGGCTACTTTGATGCTGCGCTCAGAGATTCTGTAG    | 796 |
| Qy | 721 | CAAGGGCCCTGTCTATCAAGTCTTGGGCCAATGAGAAATGGGCCCTTCATTGGTGTCCCTTAT  | 780 |
| Db | 797 | CAAGGGCCCTGTCTATCAAGTCTTGGGCCAATGAGAAATGGGCCCTTCATTGGTGTCCCTTAT  | 856 |
| Qy | 781 | GTGTCCCTCCTGTGTGGCCAAAGAAATCATCTAGTCAAGATCAAGTGA                 | 828 |
| Db | 857 | GTGTCCCTCCTGTGTGGCCAAAGAAATCATCTAGTCAAGATCAAGTGA                 | 904 |

RESULT 3  
ADN38833  
ID ADN38833 standard; cDNA; 4212 BP.

| AC | ADN38833;  | XX | SQ | Sequence 4212 BP; 1047 A; 894 C; 932 G; 1339 T; 0 U; 0 Other; | XX |
|----|--|----|----|---|----|
| XX |  |    |    |   |    |
| DT | 17-JUN-2004 (first entry)  |    |    |   |    |
| XX |  |    |    |   |    |
| DE | Cancer/angiogenesis/fibrosis-related nucleic acid, SEQ ID NO:151.        |    |    |   |    |
| XX |  |    |    |   |    |
| KW | Human; differential expression; cancer; angiogenic disorder;             |    |    |   |    |
| KW | fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis; |    |    |   |    |
| KW | inflammatory disease; autoimmune disease;                                |    |    |   |    |
| KW | retinal neovascularisation syndrome; scarring; uterine fibroid;          |    |    |   |    |
| KW | detection; diagnosis; prognosis; drug screening; drug targeting;         |    |    |   |    |
| KW | wound healing; contraception; cytostatic; cardiant; immunomodulatory;    |    |    |   |    |
| XX | vulnery; gene therapy; vaccine; gene; ss.                                |    |    |   |    |
| OS | Homo sapiens.  |    |    |   |    |
| XX |  |    |    |   |    |
| PN | WO2003042661-A2.   |    |    |   |    |
| XX |  |    |    |   |    |
| PD | 22-MAY-2003.   |    |    |   |    |
| XX |  |    |    |   |    |
| PF | 13-NOV-2002; 2002WO-US036810.  |    |    |   |    |
| XX |  |    |    |   |    |
| PR | 13-NOV-2001; 2001US-0350666P.  |    |    |   |    |
| PR | 21-NOV-2001; 2001US-0332464P.  |    |    |   |    |



QY 661 GCTGCTACCTGGGCTGTGTATGCTTTGGCTTACCTTGTGATGCTGCTCAGAGATTCCTGAG 720  
Db |||||  
QY 658 GCTGCTACCTGGGCTGTGTATGCTTTGGCTTACCTTGTGATGCTGCTCAGAGATTCCTGAG 717  
Db |||||  
QY 721 CAAGGCCCTGTCAATCAAGTCTGGCCCAATGAGAAATGAGAAATGGGCTTCATTTGGTGTCCCTAT 780  
Db |||||  
QY 718 CAAGGCCCTGTCAATCAATCTGGCCCAAGAGAAATGGGCTTCATTTGGTGTCCCTAT 777  
Db |||||  
QY 781 GTGTCCCTCCTGTGTGCAACAAGAAATCATCAGTCAAGATCAG 825  
Db |||||  
778 GTGTCCCTCCTGTGTGCAACAAGAAATCATCAGTCAAGATCAG 822

RESULT 5  
ABS54148  
ID ABS54148 standard; cDNA; 4175 BP.  
XX  
AC ABS54148;  
XX  
DT 25-NOV-2002 (first entry)  
XX  
DE Mouse cDNA encoding a liver tumour marker protein, CRG-L1.  
XX  
KW Mouse; ss; gene; liver cancer; liver tumour; CRG-L1;  
KW hepatocellular cancer.  
XX  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers  
FT CDS 35..862  
FT /\*cag= a  
FT /product= "Liver tumour marker protein, CRG-L1"  
XX  
XX US2002115094-A1.  
XX  
XX 22-AUG-2002.  
XX  
XX 14-DEC-2001; 2001US-00017410.  
XX  
XX 14-DEC-2000; 2000US-0255674P.  
XX  
XX (FARN/) FARNHAM P J.  
XX (GRAV/) GRAVEEL C R.  
XX  
XX Farnham PJ, Graveel CR;  
XX  
XX WPI; 2002-706409/76.  
XX P-PSDB; ABG32880.  
XX  
XX Novel polypeptide designated as CRG-L1, useful as diagnostic marker for  
XX liver cancer, is differentially expressed in liver tumors relative to  
XX normal liver tissues.  
XX  
XX Claim 2; Page 4-7; 11pp; English.  
XX  
XX The invention relates to a polypeptide designated as CRG-L1, which is  
XX differentially expressed in liver tumours relative to normal expression  
XX in normal liver tissues, designated CRG-L1. Also included are the  
XX encoding polynucleotides (in the case of the human sequence, mapping to  
XX chromosome 9p), expression constructs, host cells, anti-CRG-L1  
XX antibodies, identifying modulators of CRG-L1, and the use of the CRG-L1  
XX sequence in the diagnosis of hepatocellular cancer in tumour cells from a  
XX liver of a human or non-human animal. The CRG-L1 protein and  
XX polynucleotide are useful as diagnostic markers for a liver cancer in  
XX humans and non-human animals, and as a system for assessing putative  
XX therapeutic agents. The present sequence encodes mouse CRG-L1  
XX  
SQ Sequence 4175 BP; 962 A; 971 C; 1012 G; 1230 T; 0 U; 0 Other;  
Query Match 80.3%; Score 664.8; DB 6; Length 4175;  
Best Local Similarity 87.7%; Pred. No. 9.5e-192;  
Matches 726; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 1 ATGGGCGCCCGCACTGGTGGGACCAGCTGAGCTGGTAGCTGGAGGTGGAGTGTGTC 60  
Db |||||  
QY 35 ATGGGCGCCCGCACTGGTGGGACCAGCTGCGGGCTGGCAGTTCGGAGGTGATTGTGTC 94  
Db |||||  
QY 61 GAGGACAACTACACCATCTGGTGGTCTATCGCCGAGTCTTACAAACAGATCAGAAATGTC 120  
Db |||||  
QY 95 GAGGACAACTACACCATCTGGTGGTCTATCGCCGAGTCTTACAAACAGATCAGAAATGTC 154  
Db |||||  
QY 121 TTAATTTTCAATTTTACCGCCCATCTGCATGTGCTTGTGTGATGATGATGCAACATGCTTG 180  
Db |||||  
QY 155 TTGTTTTTCAATTTTACCTCCCATCTGCATGTGCTTGTTCGCCAGTACGAACTGCTTC 214  
Db |||||  
QY 181 AACAGTCAATCTACTTAATCTGGACTCTTTTGGTGTGTAGTGGAAATGGATCCGTCTAC 240  
Db |||||  
QY 215 AACAGCGGCATCTACTTAATATGACGCTCTAGTGTGTAGTGGGAAATGGATCTGTCTAC 274  
Db |||||  
QY 241 TTCCATTTTACCCCTTAGTCTTGGGTTCAGATGCTTGCATGAACTTGCAGTCTCTTGGGT 300  
Db |||||  
QY 275 TTCCATGCAACGCTGAGTTTCTGGGTTCAGATGCTTGCATGAACTTGCAGTCTCTTGGGT 334  
Db |||||  
QY 301 CTGATGTGTCTTTGGCCATGTGTTCCCGAAGAGGTATCTACCAAGATCTTTTCGAAT 360  
Db |||||  
QY 335 CTGATGTGTCTTTGGCCATGTGTTCCCGAAGAGGTATTTACCAAGATCTTTTCGAAT 394  
Db |||||  
QY 361 GACAGGGTAGGTTCAAGGTGGTGTGCTGCTGCTGCGGTTCAGACGTGCTGCGCA 420  
Db |||||  
QY 395 GACAGGGCAGGTTCAAGGCAGTGGTGTGCTGCTGCTGCAATTAACAACGTGCTGGCG 454  
Db |||||  
QY 421 TTTGTGAAGCTGCCATCAACAACATCTCTGATGACCTGGGAGTCTCTTGCACTGCA 480  
Db |||||  
QY 455 TTTATCAAGCCCGCCATCAACAATTTTCCCTGATGATTTCTGGGACTTTCCATGCACTGCG 514  
Db |||||  
QY 481 CTGCTCATCGCAGAGCTAAAGAGGTGTGAACAATGCTGTGTTTAAGCTGGGCTCTTC 540  
Db |||||  
QY 515 CTGCTTGTGCAGAGCTGAAGAGGTGTGAACAATGCTGTGTTTAAGCTGGGCTCTTC 574  
Db |||||  
QY 541 TCGGGCTCTGGTGGACCTGGGCTGTTCTGTGGATCAGTGAACCGAGTCTTTCGCGAG 600  
Db |||||  
QY 575 TCTGGCTCTGGTGGACTCTGGCTCTCTTCTGTGGATCAGCGACCAAGCCTTCTGTGAG 634  
Db |||||  
QY 601 CTGCTGTCACTTCAACTTCCCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
Db |||||  
QY 635 CTGCT 694  
Db |||||  
QY 661 GCTGCTACCTGGGCTGTGTATGCTTTGGCTTACCTTGTGATGCTGCTCAGAGATTCCTGAG 720  
Db |||||  
QY 695 GCTTGTGCTACCTGGGCTGTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTG 754  
Db |||||  
QY 721 CAAGGCCCTGTCAATCAAGTCTGGCCCAATGAGAAATGGGCTTCATTTGGTGTCCCTAT 780  
Db |||||  
QY 755 CAAGGTCCAGTCAATCAGATTTCTGGCCCAAGAGAAATGGGCTTTTATTTGGTGTCCCTAT 814  
Db |||||  
QY 781 GTGTCCCTCCTGTGTGCAACAAGAAATCATCAGTCAAGATCAGTCA 828  
Db |||||  
815 GTGTCCCTTCTGTGTGCAACAAGAGTGGCCAGATCAGTCA 862

RESULT 6  
ADI21948  
ID ADI21948 standard; cDNA; 1623 BP.  
XX  
AC ADI21948;  
XX  
DT 15-APR-2004 (first entry)  
XX  
DE Novel human protein cDNA #207.  
XX

XX forensic; nutritional source; damaged tissue; diseased tissue;  
XX myeloid cell disorder; lymphoid cell disorder;  
XX bone cartilage tissue growth; tendon tissue growth;  
XX ligament tissue growth; nerve tissue growth; regeneration; wound healing;  
XX tissue repair; tissue replacement; burn; incision; ulcer; cancer; human;

KW ss; gene.  
XX OS Homo sapiens.  
XX PN WO2003025148-A2.  
XX PD 27-MAR-2003.  
XX PF 19-SEP-2002; 2002WO-US029964.  
XX PR 19-SEP-2001; 2001US-0323739P.  
XX PR 13-SEP-2002; 2002US-00323739.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang D;  
PI Haley-Vicente D;  
XX DR WPI: 2003-354603/33.  
XX DR P-PSDB; ADI21232.  
XX PT New polynucleotides and secreted proteins, useful for treating myeloid or  
PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve  
PT tissue growth or regeneration, in wound healing, and in tissue repair and  
PT replacement.  
XX PS Claim 1; SEQ ID NO 207; 156pp; English.  
XX CC The invention relates to an isolated polynucleotide encoding a  
CC polypeptide with biological activity. The polynucleotides and  
CC polypeptides are useful in diagnostics, forensics, gene mapping,  
CC identification of mutations responsible for genetic disorders and other  
CC traits, to assess biodiversity, as nutritional sources or supplements.  
CC The polynucleotides may also be used as molecular weight markers,  
CC chromosome markers or map related gene positions, or as an antigen to  
CC raise anti-DNA antibodies or elicit immune response. The polypeptides are  
CC useful for raising antibodies, as markers for tissues in which the  
CC corresponding polypeptide is expressed, for re-engineering damaged or  
CC diseased tissues, for treating myeloid or lymphoid cell disorders, in  
CC bone cartilage, tendon, ligament and/or nerve tissue growth or  
CC regeneration, in wound healing, in tissue repair and replacement, in  
CC healing of burns, incisions and ulcers, and in treating cancer. The  
CC present sequence represents cDNA encoding a novel human protein.  
XX SQ Sequence 1623 BP; 305 A; 457 C; 382 G; 479 T; 0 U; 0 Other;  
Query Match 58.7%; Score 486.2; DB 10; Length 1623;  
Best Local Similarity 97.4%; Pred. No. 2e-137;  
Matches 494; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 322 TGGTCCCGAGAGGTATCTACCAAGATCTTTCGGAATGACAGGGGTAGGTTCAAGGTG 381  
DB 920 TTGTTTCGTCAAAAGGTATCTACCAAGATCTTTCGGAATGACAGGGGTAGGTTCAAGGTG 979  
QY 382 GTGGTCAGTCTCCTGCTGGGTTTACGAGCTGCTGCGATTTCTCAAGCTGCCATCAAC 441  
DB 980 GTGGTCTGTCTCTGCTGAGTTATGACGTGCTGCGATTTCTCAAGCTGCCATCAAC 1039  
QY 442 AACATCTCTCTGATGACCCCTGGAGTTCCTTGCACTGCACTGCTCATCGCAGAGCTAAAG 501  
DB 1040 AACATCTCTCTGATGACCCCTGGAGTTCCTTGCGCTGCACTGCTCATCAGAGCTAAAG 1099  
QY 502 AGGTGTGACCAACATCGGTGTGTTTAAGCTGGGCTCTTCTCGGGCTCTGGTGACCCCTG 561  
DB 1100 AGGTGTGACCAACATCGGTGTGTTTAAGCTGGGCTCTTCTCGGGCTCTGGTGACCCCTG 1159  
QY 562 GCCTGTGTCGTGGATCAGTGACCGAGCTTTCTGGGAGCTGTGTGTCATCTTCAACTTC 621  
DB 1160 GCCTGTGTCGTGGATCAGTGACCGAGCTTTCTGGGAGCTGTGTGTCATCTTCAACTTC 1219  
QY 622 CCCTACTGTCAGTGCATGTGGCAATCCTCATCTGCGCTTCTGCTGCTACCTGGGCTGTGTA 681

DB 1220 CCTACCTGCATGCTGCGACATCTCATCTGCTTGCCTACCTGCGGTGTGTA 1279  
QY 682 TGGTTTGCTTACTTTGATGCTGCTCAGAGATTCCTGAGCAAGGCCCTGTCTCATCAAGTTC 741  
DB 1280 TGGTTTGCTTACTTTGATGCTGCTCAGAGATTCCTGAGCAAGGCCCTGTCTCATCAAGTTC 1339  
QY 742 TGGSCCAATGAGAAATGGGCTTCAATGGTGTCCTTATGTGTCCTCTCTGTGTGCAAC 801  
DB 1340 TGGSCCAATGAGAAATGGGCTTCAATGGTGTCCTTATGTGTCCTCTCTGTGTGCAAC 1399  
QY 802 AAGAAATCATCATGCTCAAGATCAGTGA 828  
DB 1400 AAGAAATCATCATGCTCAAGATCAGTGA 1426  
RESULT 7  
ADL06640  
ID ADL06640 standard; cDNA; 1527 BP.  
XX AC ADL06640;  
XX DT 06-MAY-2004 (first entry)  
XX DE Human 3T3 cell conversion promoter PP11646 cDNA.  
XX KW 3T3 cell conversion; promoter; human; gene; ss.  
XX OS Homo sapiens.  
XX PN CN1403477-A.  
XX PD 19-MAR-2003.  
XX PF 12-SEP-2001; 2001CN-00126725.  
XX PR 12-SEP-2001; 2001CN-00126725.  
XX PA (SHAN-) SHANGHAI XINSHIJI GENE TECHN DEV CO LTD.  
XX PI Gu J, Yang S;  
XX WPI: 2003-494226/47.  
XX P-PSDB; ADL06641.  
XX PT Human protein with function of promoting 3T3 cell conversion and its  
XX coding sequence.  
XX Example 1; SEQ ID NO 1; 41pp; Chinese.  
XX CC This invention describes a novel human protein with 3T3 cell conversion  
XX promoting function, polynucleotides encoding the polypeptide and the  
XX recombinant process of producing the polypeptide. The present invention  
XX also discloses the agonist resisting the polypeptide and its treatment  
XX effect. The present invention also discloses the application of the  
XX polynucleotides encoding the human protein with 3T3 cell conversion  
XX promoting function.  
XX SQ Sequence 1527 BP; 376 A; 357 C; 383 G; 411 T; 0 U; 0 Other;  
Query Match 50.1%; Score 414.8; DB 10; Length 1527;  
Best Local Similarity 82.4%; Pred. No. 1.3e-115;  
Matches 532; Conservative 0; Mismatches 17; Indels 97; Gaps 1;  
QY 108 GATCAGCAATGTCTTATTTTTCATTTTACGGCCATCTGCATGCTGTTGATGAGTA 167  
DB 178 GATCAGCAATGTCTTATTTTTCATTTTACGGCCATCTGCATGCTGTTGATGAGTA 237  
QY 168 TGCAACATGCTTGAACAGATGACATCTACTTAATCTGGACTCTTTTGGTTGTAGTGGAAAT 227  
DB 238 TGCAACATGCTTCAACAGATGGCATCTACTTAATCTGGACTCTTTTGGTTGTAGTGGAAAT 297  
QY 228 TGATCCGCTACTTCCATTTTACCTTAGTTTCTTGGGTGAGATGCTTGTGAACTTGC 287

298 TGGATCCGCTACTTCCATGCAACCCCTTAGTTTCTTGGTCAGATGCTTGATGAATGTC 357  
288 AGTCCTTTGGGTTCTGATGTGTGCTTTGGCCATGTGGTTCCCGAAGGTATCTACCAA 347  
358 AGTCCTTTGGGTTCTGATGTGTGCTTTGGCCATGTGGTTCCCGAAGGTATCTACCAA 417  
348 GATCTTTCCGAATGACAGGGGTAGGTTCAAGGTGTGTGCTGATGTGCTGCGGTTC 407  
418 GATCTTTCCGAATGACCGGGGTAGGTTCAAGGTGTGTGCTGATGTGCTGCGGTTC 477  
408 GACGTGCTGCTGCTTGTCAAGCCTGCCATCAACAACTCTCTGTGATGACCCCTGGAGT 467  
478 GACGTGCTGCTGCTTGTCAAGCCTGCCATCAACAACTCTCTGTGATGACCCCTGGAGT 537  
468 TCCTTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 501  
538 TCCTTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597  
502 ----- 501  
598 ACACAGGAAGGTGGCCAGGAGGTGGAGACAAGGTCTGACATGATGATCTCTCTG 657  
502 -----AGTGTGACAACTATGCGTGTGTTTAAAGTGGGCTCTTCTCGGGCTCT 550  
658 GACCCCGTGCAGGTGTGACAACTATGCGTGTGTTTAAAGTGGGCTCTTCTCGGGCTCT 717  
551 GGTGGACCTGGCCCTGTTCTGCTGATCAGTGCAGGCTTCTGCGAGTGTGCTGCT 610  
718 GGTGGACCTGGCCCTGTTCTGCTGATCAGTGCAGGCTTCTGCGAGTGTGCTGCT 777  
611 CCTTCAACTTCCCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 656  
778 CCTTCAACTTCCCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 823

## RESULT 8

ADL06642  
ID ADL06642 standard; DNA; 1527 BP.  
XX  
AC ADL06642;  
XX  
XX 06-MAY-2004 (first entry)  
XX Human 3T3 cell conversion promoter Pp11646 DNA.  
XX 3T3 cell conversion; promoter; human; gene; ds.  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX CDS 218..640  
XX /\*tag= a  
XX /product= "Pp11646"  
XX  
XX CN1403477-A.  
XX  
XX 19-MAR-2003.  
XX  
XX 12-SEP-2001; 2001CN-00126725.  
XX  
XX 12-SEP-2001; 2001CN-00126725.  
XX  
XX (SHAN-) SHANGHAI XINSHIJI GENE TECHN DEV CO LTD.  
XX Gu J, Yang S;  
XX  
XX WPI; 2003-494226/47.  
XX P-PSDB; ADL06641.  
XX  
XX Human protein with function of promoting 3T3 cell conversion and its  
XX coding sequence.  
XX  
XX Example 1; SEQ ID NO 3; 41pp; Chinese.

XX This invention describes a novel human protein with 3T3 cell conversion  
CC promoting function, polynucleotides encoding the polypeptide and the  
CC recombinant process of producing the polypeptide. The present invention  
CC also discloses the agonist resisting the polypeptide and its treatment  
CC effect. The present invention also discloses the application of the  
CC polynucleotides encoding the human protein with 3T3 cell conversion  
CC promoting function.  
XX

Sequence 1527 BP; 376 A; 357 C; 383 G; 411 T; 0 U; 0 Other;

Query Match 50.1%; Score 414.8; DB 10; Length 1527;

Best Local Similarity 82.4%; Pred. No. 1.3e-115;

Matches 532; Conservative 0; Mismatches 17; Indels 97; Gaps 1;

QY 108 GATCAGCAATGCTCTATTTTTCATTTTACCCGCCATCTGCATGTGCTTGTGATGAGTA 167  
DB 178 GATCAGCAATGCTCTATTTTTCATTTTACCCGCCATCTGCATGTGCTTGTGATGAGTA 237  
QY 168 TGCACATGCTTGACAGTGCATCTACTTAATCTGCATCTTTTGGTGTGATGGGAAT 227  
DB 238 TGCACATGCTTGACAGTGCATCTACTTAATCTGCATCTTTTGGTGTGATGGGAAT 297  
QY 228 TGGATCGTCTACTTCCATTTTACCCCTTACTTCTTGGTTCAGATGCTTGTGATGAGTGC 287  
DB 298 TGGATCGTCTACTTCCATGCAACCCCTTAGTTTCTTGGTTCAGATGCTTGTGATGAGTGC 357  
QY 288 AGTCCTTTGGGTTCTGATGTGTGCTTTGGCCATGTGCTTCCCGAAGGTATCTACCAA 347  
DB 358 AGTCCTTTGGGTTCTGATGTGTGCTTTGGCCATGTGCTTCCCGAAGGTATCTACCAA 417  
QY 348 GATCTTTCCGAATGACAGGGGTAGGTTCAAGGTGTGCTGATGCTGCTGCTGCTGCTGCT 407  
DB 418 GATCTTTCCGAATGACCGGGGTAGGTTCAAGGTGTGCTGATGCTGCTGCTGCTGCTGCT 477  
QY 408 GACGTGCTGCTGCTTGTCAAGCCTGCCATCAACAACTCTCTGTGATGACCCCTGGAGT 467  
DB 478 GACGTGCTGCTGCTTGTCAAGCCTGCCATCAACAACTCTCTGTGATGACCCCTGGAGT 537  
QY 468 TCCTTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 501  
DB 538 TCCTTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597  
QY 502 ----- 501  
DB 598 ACACAGGAAGGTGGCCAGGAGGTGGAGACAAGGTCTGACATGATGATCTCTCTG 657  
QY 502 -----AGTGTGACAACTATGCGTGTGTTTAAAGTGGGCTCTTCTCGGGCTCT 550  
DB 658 GACCCCGTGCAGGTGTGACAACTATGCGTGTGTTTAAAGTGGGCTCTTCTCGGGCTCT 717  
QY 551 GGTGGACCTGGCCCTGTTCTGCTGATCAGTGCAGGCTTCTGCGAGTGTGCTGCTGCT 610  
DB 718 GGTGGACCTGGCCCTGTTCTGCTGATCAGTGCAGGCTTCTGCGAGTGTGCTGCTGCT 777  
QY 611 CCTTCAACTTCCCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 656  
DB 778 CCTTCAACTTCCCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 823

## RESULT 9

ADL21468  
ID ADL21468 standard; cDNA; 487 BP.

XX  
XX AC ADL21468;

XX  
XX AC ADL21468;

XX  
XX 15-APR-2004 (first entry)

XX Novel human expressed sequence tag, EST #167.  
XX forensic; nutritional source; damaged tissue; diseased tissue;  
XX myeloid cell disorder; lymphoid cell disorder;  
XX bone cartilage tissue growth; tendon tissue growth;

KW ligament tissue growth; nerve tissue growth; regeneration; wound healing;  
KW tissue repair; tissue replacement; burn; incision; ulcer; cancer; human;  
ss; gene; EST; expressed sequence tag.

XX Homo sapiens.

OS  
XX WO2003025148-A2.

XX  
XX 27-MAR-2003.

XX  
XX 19-SEP-2002; 2002WO-US029964.

XX  
XX 19-SEP-2001; 2001US-0323739P.

PR  
PR 13-SEP-2002; 2002US-00323739.

XX  
XX (HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang D;  
PI Haley-Vicente D;

XX  
XX WPI; 2003-354603/33.

DR  
DR P-PSDB; ADI21688.

XX New polynucleotides and secreted proteins, useful for treating myeloid or  
PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve  
PT tissue growth or regeneration, in wound healing, and in tissue repair and  
PT replacement.

XX Example 2; SEQ ID NO 719; 156pp; English.

XX The invention relates to an isolated polynucleotide encoding a  
CC polypeptide with biological activity. The polynucleotides and  
CC polypeptides are useful in diagnostics, forensics, gene mapping,  
CC identification of mutations responsible for genetic disorders and other  
CC traits, to assess biodiversity, as nutritional sources or supplements.

CC The polynucleotides may also be used as molecular weight markers.  
CC chromosome markers or map related gene positions, or as an antigen to  
CC raise anti-DNA antibodies or elicit immune response. The polypeptides are  
CC useful for raising antibodies, as markers for tissues in which the  
CC corresponding polypeptide is expressed, for re-engineering damaged or  
CC diseased tissues, for treating myeloid or lymphoid cell disorders, in  
CC bone cartilage, tendon, ligament and/or nerve tissue growth or  
CC regeneration, in wound healing, in tissue repair and replacement, in  
CC healing of burns, incisions and ulcers, and in treating cancer. The  
CC present sequence represents a novel human expressed sequence tag, EST.

XX Sequence 487 BP; 81 A; 138 C; 131 G; 134 T; 0 U; 3 Other;

PS Query Match 37.8%; Score 313; DB 10; Length 487;

XX Best Local Similarity 97.0%; Pred. No. 8.5e-85;

XX Matches 319; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ATGGGGCCCGCCGACATCGTGGGACACGCTGCGAGTGTAGCTCGGAGGTGACTGGTGC 60

DB 159 ATGGGGCCCGCCGACATCGTGGGACACGCTGCGAGTGTAGCTCGGAGGTGACTGGTGC 218

QY 61 GAGGACAACTACACCATCGTGGTGTCTATCGCGAGTTCTTACACGATCAGCAATGTC 120

DB 219 GAGGACAACTACACCATCGTGGTGTCTATCGCGAGTTCTTACACGATCAGCAATGTC 278

QY 121 TTAATTTTCAATTTTACCGCCCATCTCATGCTGCTTGTTCATGATGATGCAATGCTTC 180

DB 279 TTAATTTTCAATTTTACCGCCCATCTCATGCTGCTTGTTCATGATGATGCAATGCTTC 338

QY 181 AACAGTGACATCTACTTAATCTGGACTCTTTGGTTGTAGTGGAAATGGATCCGCTAC 240

DB 339 AACAGTGACATCTACTTAATCTGGACTCTTTGGTTGTAGTGGAAATGGATCCGCTAC 398

QY 241 TTCAGTTTACCCCTTAGTTTCTTGGGTCAGATGCTTGATGAATTCAGTCCCTTGGGTT 300

DB 399 TTCAGTTTACCCCTTAGTTTCTTGGGTCAGATGCTTGATGAATTCAGTCCCTTGGGTT 458

QY 301 CTGATGTGTGCTTTGGCCATGTGGTTCCTCC 329

DB 459 CTGATGTGTGCTTTAGCCATGTGGTTCCTCC 487

RESULT 10

AAKS1566

ID AAKS1566 standard; cDNA; 1215 BP.

XX

XX AAKS1566;

XX 06-NOV-2001 (first entry)

XX Human polynucleotide SEQ ID NO 111.

XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation; ss.

XX Homo sapiens.

XX WO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.

XX 27-APR-2000; 2000US-00560875.

XX 20-JUN-2000; 2000US-00598075.

XX 19-JUL-2000; 2000US-00620325.

XX 01-SEP-2000; 2000US-00654936.

XX 15-SEP-2000; 2000US-00663561.

XX 20-OCT-2000; 2000US-00693325.

XX 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX Xue AJ, Yang Y, Wehrman T, Goodrich R;

XX WPI; 2001-476283/51.

XX P-PSDB; AAM78433.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful  
PT in diagnosis and gene therapy.

XX Claim 1; Page 750-751; 6221pp; English.

XX The invention relates to polynucleotides (AAKS1456-AAKS3435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAKS2581), 2111  
CC (AAKS2582) and 3666 (AAM80020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication

XX Sequence 1215 BP; 213 A; 374 C; 309 G; 319 T; 0 U; 0 Other;

XX Query Match 22.3%; Score 185; DB 4; Length 1215;

XX Best Local Similarity 53.3%; Pred. No. 1.7e-45;

XX Matches 414; Conservative 0; Mismatches 360; Indels 3; Gaps 1;

QY 40 AGTCGGAGGTGACCTGGTGGCGAGGCAACTACCATCGTCTGCTATCGCGAGTTC 99

Db 80 AGTCCGAGGTGGACTGGTGTGAGAGCAACTCCAGTACTCGAGCTGGTGGCCAGTTC 139  
 Qy 100 TACAACAGGATCAGAAATGCTTATTTTATTTTACCGCCCATCTGATGCTGTGTTT 159  
 Db 140 TACAACAGGATCAGAAATGCTTATTTTATTTTACCGCCCATCTGATGCTGTGTTT 199  
 Qy 160 GATGAGTATGCAACATGCTTGAACAGTGCATCTACTTAACTTGAACATCTTTTGGTCTA 219  
 Db 200 CACCGTATGCCAGAGCGCTCCGCTACATTTACGTGTGCTGGTCTCTTTCATGATC 259  
 Qy 220 GTGGAAATGGATCCGCTACTTCCATTTTACCTTAGTTTCTTGGTCCAGATGCTGTAT 279  
 Db 260 ATAGSCCTGTTCTCAATGATTTCCACATGAGCTCAGCTTCTTGGGCCAGCTGCTGAC 319  
 Qy 280 GAACCTTGAGTCTTGGTGTCTGATGCTGCTTGGCCATGCTGCTTCCCGAGAGGAT 339  
 Db 320 GAGATCGCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAT 379  
 Qy 340 CTACCAAGATCTTTCGGAATGACAGGGGTAGTTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 399  
 Db 380 TTCCTCTCTCTTGGGGGAAACAGGTCCTCAGTTTCAATCCGCTGCTTCAATCACCCT 439  
 Qy 400 GCGGTTACAGCTGCTGGCATTTTGTCAAGCTTGGCATCAACATCTCTCTGATGAC 459  
 Db 440 GTGTCAGCACCTTCTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 499  
 Qy 460 CTGGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 519  
 Db 500 ATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 559  
 Qy 520 GTGTTTAAAGCTGGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 579  
 Db 560 CTGCGCACCTTATGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 619  
 Qy 580 AGTGACCGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 639  
 Db 620 AGTGACCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 679  
 Qy 640 TGGCAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 699  
 Db 680 TGGCAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 739  
 Qy 700 GCTGCTCAGAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 759  
 Db 740 GCAACTATGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 799  
 Qy 760 GCCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 816  
 Db 800 CC---CGTGGGCTGCCCTAGCTGGAATCGGGGTGATGACAGGACTGCTGAGAC 853

RESULT 11

ID AAK52550 standard; cDNA; 1631 BP.  
 XX  
 AC AAK52550;  
 XX  
 DE 06-NOV-2001 (first entry)  
 XX Human polynucleotide SEQ ID NO 2079.  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620325.  
 PR 01-SEP-2000; 2000US-00654936.  
 PR 15-SEP-2000; 2000US-00683561.  
 PR 20-OCT-2000; 2000US-00693325.  
 PR 30-NOV-2000; 2000US-00728422.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
 PI Ma Y, Zhao QA, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX WPI; 2001-476283/51.  
 DR P-PSDB; AAM79417.  
 XX  
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 PT in diagnosis and gene therapy.  
 XX  
 PS Claim 1; Page 4462; 6221pp; English.  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication

Sequence 1631 BP; 247 A; 399 C; 344 G; 350 T; 0 U; 291 Other;

Query Match 22.3%; Score 185; DB 4; Length 1631;  
Best Local Similarity 53.3%; Pred. No. 2e-45;

Matches 414; Conservative 0; Mismatches 360; Indels 3; Gaps 1;

Qy 40 AGCTCGAGGTGGACTGGTGTGAGAGCAACTCCAGTACTCGAGCTGGTGGCCAGTTC 99  
 Db 80 AGCTCGAGGTGGACTGGTGTGAGAGCAACTCCAGTACTCGAGCTGGTGGCCAGTTC 139  
 Qy 100 TACAACAGGATCAGAAATGCTTATTTTATTTTACCGCCCATCTGATGCTGTGTTT 159  
 Db 140 TACAACAGGATCAGAAATGCTTATTTTATTTTACCGCCCATCTGATGCTGTGTTT 199  
 Qy 160 GATGAGTATGCAACATGCTTGAACAGTGCATCTACTTAACTTGAACATCTTTTGGTCTA 219  
 Db 200 CACCGTATGCCAGAGCGCTCCGCTACATTTACGTGTGCTGGTCTCTTTCATGATC 259  
 Qy 220 GTGGAAATGGATCCGCTACTTCCATTTTACCTTAGTTTCTTGGTCCAGATGCTGTAT 279  
 Db 260 ATAGSCCTGTTCTCAATGATTTCCACATGAGCTCAGCTTCTTGGGCCAGCTGCTGAC 319  
 Qy 280 GAACCTTGAGTCTTGGTGTCTGATGCTGCTTGGCCATGCTGCTTCCCGAGAGGAT 339  
 Db 320 GAGATCGCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAT 379  
 Qy 340 CTACCAAGATCTTTCGGAATGACAGGGGTAGTTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 399  
 Db 380 TTCCTCTCTCTTGGGGGAAACAGGTCCTCAGTTTCAATCCGCTGCTTCAATCACCCT 439  
 Qy 400 GCGGTTACAGCTGCTGGCATTTTGTCAAGCTTGGCATCAACATCTCTCTGATGAC 459  
 Db 440 GTGTCAGCACCTTCTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 499  
 Qy 460 CTGGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 519

Db 500 ATTGCGCTGCACATCTCTACATCGTGTGCCAGAGTACAGAGACCAACATAGGAG 559  
QY 520 GTGTTTAAGCTGGGCTCTTCTCGGCGCTCTGTGGACCTTGGCCCTGTCTGTGATC 579  
Db 560 CTTGCGGACATGATGAGGTCTCGTGGTCTTATGGCTGTGTCTGACACAGCTGATC 619  
QY 580 AGTGACCGAGCTTCTGCGAGCTGCTGCATCCTTCAACTTCCCTACCTGCACTGATG 639  
Db 620 AGTGACCGCTGTGCTGAGCTTCTGCGAGAGATTCATTTCTTATCTGCAAGATC 679  
QY 640 TGGCACATCTCATCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 699  
Db 680 TGGCATGTGCTCATCAGCATCCTTCCCTTATGCGATGTCTACCATGGCTTGGTGTAT 739  
QY 700 GCTGCTCAGAGATCTCTGAGCAAGGCTGTGATCAAGTCTTGGCCCAATGAGAAATGG 759  
Db 740 GCCAACTATGAGATGCCAGGTGAACCTCAAAAGTCCGCTACTGGCTCGGACAGTTGG 799  
QY 760 GCCTTCAATGCTGCTCCCTATGCTCCTCTGTCGCAACCAAGAAATCATCAGTC 816  
Db 800 CC---CGTGGGCTGCCCTACGTGGAATCCGGGTGATGACAAAGACTCTCTGAGAC 853

## RESULT 12

AAH48639  
ID AAH48639 standard; DNA; 792 BP.

XX AAH48639;

XX 21-SBP-2001 (first entry)

XX Human ceramidase K1 DNA.

XX Ceramidase; human; K1; antiproliferative; anticancer; anti-eczema;  
KW antipsoriasis; dermatological; ceramide; sphingosine; treatment; ds;  
KW altered cell proliferation; cancer; neurodermatitis; eczema; cosmetic;  
KW permeability barrier; psoriasis; ichthyosis; lamellar ichthyosis ICR2.  
XX Homo sapiens.

XX WO200155408-A1.

XX 02-AUG-2001.

XX 27-JAN-2001; 2001WO-EP000900.

XX 27-JAN-2000; 2000DE-01003293.

PR 09-MAR-2000; 2000DE-01011392.

XX (MEMO-) MEMOREC MEDICAL MOLECULAR RES COLOGNE ST.

XX Hofmann K, Conradt M;

XX WPI; 2001-483256/52.

DR P-PSDB; AAB86363.

XX New ceramidase containing specific structural motif, useful for  
PT diagnosis, prevention and treatment of ceramidase defects, e.g.  
PT ichthyosis, also in cosmetics.

XX Claim 3; Page 16; 31pp; German.

XX This invention describes novel human ceramidase (I) containing a specific  
CC structural motif which has antiproliferative, anticancer, anti-eczema,  
CC antipsoriasis and dermatological activity. (I) cleave ceramide, resulting  
CC in formation of sphingosine. (II), or the nucleic acid (II) encoding it,  
CC are used for diagnosis, prevention or treatment of diseases associated  
CC with ceramide defects, particularly altered cell proliferation (cancer)  
CC or altered ceramide layers on the skin (neurodermatitis, eczema,  
CC psoriasis), also for targeted modification of the permeability barrier by  
CC ceramidase or its activators, e.g. for transcutaneous delivery of  
CC substances. Specifically they are used for diagnosis of ichthyosis,

CC particularly lamellar ichthyosis ICR2. (I) can also be used in cosmetics.  
CC This sequence encodes human ceramidase K1

XX Sequence 792 BP; 145 A; 241 C; 200 G; 206 T; 0 U; 0 Other;

QY Query Match 22.1%; Score 183.4; DB 4; Length 792;

XX Best Local Similarity 53.3%; Pred. No. 4.2e-45;

XX Matches 410; Conservative 0; Mismatches 356; Indels 3; Gaps 1;

QY 40 AGTCGGAGGTGACTGTGTGGAGGACAACTACCATCGTGTGCTGTATCGCGAGTTTC 99

Db 25 AGTCGGAGGTGACTGTGTGGAGGACAACTTCCAGTACTCGGAGCTGTGTGGCGAGTTTC 84

QY 100 TACAACACGATCAGCAATGCTTATTTTTCATTTTACCGCCCATCTGCATGTGCTTCTTT 159

Db 85 TACAACACGCTTCCCAATATCCCTTCTTTCATCTTGGGCCACTGATGCTCTCTGATG 144

QY 160 GATGAGTATGCAACATGCTTTGAAACAGTGCATCTACTTAATCTGGACTCTTTTGGTTGTA 219

Db 145 CACCCGATATGCCAGAACGCTCCCGCTACATTTACGTTGTCTGGGTCTCTTTCATGATC 204

QY 220 GTGGGAATTTGGATCCGCTTACTTCCATTTTACCTTAGTTTCTTGGGTGAGATGCTTGTAT 279

Db 205 ATAGGGCTGTCTTCCATGATTTTCCATGACGCTCAGCTTCTCTGGGCCAGCTGCTGGAC 264

QY 280 GAACCTTGCACTGCTTGGGGTCTGATGTGTGCTTTTGGCCATGTGTTTCCCGCAAGATAT 339

Db 265 GAGATGCCATCTCTGGCTCTCTGGGAGTGGCTATAGCATATGGATGCCCCGCTGTAT 324

QY 340 CTACCAAGATCTTTTCGGAATGACAGGGGTAGGTTTCAAGGTGTGTGTGCTGCTGTCT 399

Db 325 TTCCCTCTCTTCTTGGGGGAAACAGTCCAGTTTATCGCTGTGTTCTTTCATCACCCT 384

QY 400 GCGGTTACGACGTGCTTGGCATTTGTCAAGCTTGCATCAACATCTCTCTGATGACC 459

Db 385 GTGGTCAGCACCTTCTCTGCTTCTTCTGGGCCCAACGCTCAACGCTTCAACGCTTCAACG 444

QY 460 CTGGGAGTTCCTTGGCACTGCACTGCTCATCGCAGAGCTAAAGAGGTGTGCAACATCGCT 519

Db 445 ATTGCCCTGCACATTTCTTACATCGTGTGCCAGAGTACAGGAAGACCACATTAAGAG 504

QY 520 GTGTTTAAGCTGGGCTCTTCTCGGCGCTCTGGTGACCTTGGCCCTGTTTCTGCTGATC 579

Db 505 CTTGCGGACCTGATTTGAGGTCTCCGTGGTCTTATGGCTGTGCTCTGACACAGCTGATC 564

QY 580 AGTGACCGAGCTTCTGCGAGCTGTGTCATCTTCAACTTCCCTTACCTGCACTGCTGATG 639

Db 565 AGTGACCGCTGTGCTTTGCACTTCTGGCAGAGATTCATTTCTTATCTGCAAGATC 624

QY 640 TGGCACATCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 699

Db 625 TGGCATGTGCTCATCAGCATCCTTCCCTTATGGCATGTGTACCCTGCTTGGTGTAT 684

QY 700 GCTGCTCAGAGATCTCTGAGCAAGGCTTGTGTCATCAAGTCTTGGGCCAATGAGAAATGG 759

Db 685 GCCAACTATGAGATGCCAGGTGAACCTTCAAAAGTCCGCTACTGCTGGCTCGGACAGTTGG 744

QY 760 GCCTTCAATGCTGCTCCCTATGCTGCTCTCTGTCGCAACCAAGAAATCATCAGTAT 808

Db 745 CC---CGTGGGCTGCCCTACGTGGAATCCGGGTGATGACAAAGGACT 790

## RESULT 13

ABL24043

ID ABL24043 standard; DNA; 537 BP.

XX ABL24043;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 23602.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW Drosophila; developmental biology; cell signalling; insecticide;





Db 3404 ACTTCTAGTGGNAATCGGAGCTGCCCGACGACGACGGCTGTGTAAGTACTGGCCAAAGA 3463  
 QY 752 AGAAATGGGCTTC 765  
 Db 3464 ACGAGTTCGAGTTC 3477

RESULT 15

ABL24042  
 ID ABL24042 standard; DNA; 4758 BP.

XX AC ABL24042;  
 XX DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 23599.  
 XX DE Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX OS Drosophila melanogaster.  
 XX PN WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX PF 23-MAR-2001; 2001WO-US009231.  
 XX PR 23-MAR-2000; 2000US-0191637P.  
 XX PR 11-JUL-2000; 2000US-00614150.  
 XX PA (PEKE ) PE CORP NY.  
 XX PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX WIPI; 2001-656860/75.  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX PS Claim 1; SEQ ID NO 23599; 21pp + Sequence Listing; English.  
 XX CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABBS72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 4758 BP; 1304 A; 981 C; 980 G; 1493 T; 0 U; 0 Other;  
 Query Match 9.1%; Score 75; DB 4; Length 4758;  
 Best Local Similarity 55.6%; Pred. No. 1.4e-11;  
 Matches 144; Conservative 0; Mismatches 115; Indels 0; Gaps 0;  
 QY 108 GATCAGCAATGCTTATTTTTCATTTTACCGCCCATCTGCATGTGCTTGTGATGAGTA 167  
 Db 3287 GTTTAGCAACTTCTGTTCTACTTCTGCGCGCGTCTTATATGCTCTTCAGGAGTA 3346  
 QY 168 TGCAACATGCTGAACAGTGCATCTACTTAATCTGGACTCTTTTGGTGTGTAGTGGAAAT 227  
 Db 3347 CGGACGGCTTTGTGACGCGCGGAATCCACGTCATCTGGGTGTGCTCATCGTGTGGCT 3406  
 QY 228 TGGATCGCTTACTTCCATTTTACCTTACTTCTTGGTTCAGATGCTTGTGACTTGC 287  
 Db 3407 GAGTTCGATGTACTTCCATTCACCTTTGAGTCTGATTTGGCCAGCTGCTGGACGAACGGC 3466  
 QY 288 AGTCCTTTGGGTTCTGTGATGTGTGCTTTGGCCATGTGCTTTGCCAGAGGTATCTACCAA 347

Db 3467 CATACTCTGGTCTTTCATGGCGGCTTTTGGCTCTTCTATCCGAGGATACTATCCAA 3526  
 QY 348 GATCTTTTCGGAATGACAGG 366  
 Db 3527 GTTCGTGAAAAACGATCGG 3545

Search completed: November 20, 2004, 21:28:48  
 Job time : 355.193 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 20:51:27 ; Search time 70.8343 Seconds  
(without alignments)  
8308.591 Million cell updates/sec

Title: US-10-017-410-3

Perfect score: 828  
Sequence: 1 atggggcccccgaactggtg.....catcaagtaagatcaagtga 828

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A-COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B-COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A-COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B-COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS-COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description         |
|------------|-------|-------------|--------|-------|---------------------|
| C 1        | 108.4 | 13.1        | 358    | 4     | US-09-621-976-11241 |
| C 2        | 49.4  | 6.0         | 7218   | 1     | Sequence 11241, A   |
| C 3        | 40.8  | 4.9         | 289    | 3     | Sequence 14, Appl   |
| C 4        | 40.8  | 4.9         | 289    | 3     | Sequence 17, Appl   |
| 5          | 40.2  | 4.9         | 1063   | 4     | Sequence 875, Appl  |
| 6          | 37.4  | 4.5         | 2514   | 3     | Sequence 3, Appl    |
| 7          | 35.8  | 4.3         | 1020   | 4     | Sequence 106, Appl  |
| 8          | 35.4  | 4.3         | 2340   | 4     | Sequence 67, Appl   |
| 9          | 35.4  | 4.3         | 2385   | 4     | Sequence 68, Appl   |
| 10         | 35.4  | 4.3         | 2538   | 4     | Sequence 65, Appl   |
| 11         | 35.4  | 4.3         | 2553   | 4     | Sequence 66, Appl   |
| 12         | 35.4  | 4.3         | 2634   | 4     | Sequence 64, Appl   |
| 13         | 35.4  | 4.3         | 4472   | 4     | Sequence 75, Appl   |
| 14         | 35.4  | 4.3         | 4766   | 4     | Sequence 73, Appl   |
| 15         | 35.2  | 4.3         | 5275   | 1     | Sequence 1, Appl    |
| 16         | 35.2  | 4.3         | 5275   | 1     | Sequence 1, Appl    |
| 17         | 35.2  | 4.3         | 5275   | 2     | Sequence 1, Appl    |
| 18         | 35.2  | 4.3         | 5275   | 2     | Sequence 1, Appl    |
| 19         | 35.2  | 4.3         | 5275   | 3     | Sequence 1, Appl    |
| 20         | 35.2  | 4.3         | 5275   | 3     | Sequence 1, Appl    |
| 21         | 35.2  | 4.3         | 5275   | 3     | Sequence 1, Appl    |
| 22         | 34.8  | 4.2         | 505    | 4     | Sequence 15639, A   |
| 23         | 34.8  | 4.2         | 1359   | 4     | Sequence 1224, Ap   |
| 24         | 34.6  | 4.2         | 505    | 4     | Sequence 15639, A   |
| 25         | 34    | 4.1         | 1513   | 1     | Sequence 2, Appl    |
| 26         | 34    | 4.1         | 1546   | 1     | Sequence 3, Appl    |
| C 27       | 34    | 4.1         | 2064   | 1     | Sequence 30, Appl   |

C 28 34 4.1 2064 1 US-08-164-839-32 Sequence 32, Appl  
C 29 34 4.1 2064 1 US-08-583-799-30 Sequence 30, Appl  
C 30 34 4.1 2064 1 US-08-583-799-32 Sequence 32, Appl  
C 31 34 4.1 2148 1 US-08-164-839-69 Sequence 69, Appl  
C 32 34 4.1 2148 1 US-08-164-839-71 Sequence 71, Appl  
C 33 34 4.1 2148 1 US-08-583-799-69 Sequence 69, Appl  
C 34 34 4.1 2148 1 US-08-583-799-71 Sequence 71, Appl  
C 35 34 4.1 3172 1 US-08-314-309A-1 Sequence 1, Appl  
C 36 33.6 4.1 2016 3 US-09-132-118-1 Sequence 1, Appl  
C 37 33.6 4.1 2137 1 US-08-444-005-16 Sequence 16, Appl  
C 38 33.6 4.1 2617 3 US-09-161-443-1 Sequence 1, Appl  
C 39 33.6 4.1 2617 4 US-09-023-655-914 Sequence 914, App  
C 40 33.6 4.1 9687 3 US-09-133-944-2 Sequence 2, Appl  
C 41 33.6 4.1 9687 3 US-09-208-827-2 Sequence 2, Appl  
C 42 33.6 4.1 9687 4 US-10-043-074-2 Sequence 2, Appl  
C 43 33.6 4.1 4403765 3 US-09-103-840A-2 Sequence 1, Appl  
C 44 33.6 4.1 4411529 3 US-09-103-840A-1 Sequence 1, Appl  
C 45 33.2 4.0 474 4 US-09-621-976-18033 Sequence 18033, A

## ALIGNMENTS

RESULT 1  
US-09-621-976-11241/c  
; Sequence 11241, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 11241  
; LENGTH: 358  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-11241

Query Match 13.1%; Score 108.4; DB 4; Length 358;  
Best Local Similarity 99.1%; Pred. No. 2.4e-23;  
Matches 109; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGGCCCCCGCACTGGTGGGACCAGCTGCAGGCTGTAGCTCGGAGTGGACTGGTGC 60  
Db 111 ATGGGCCCCCGCACTGGTGGGACCAGCTGCAGGCTGTAGCTCGGAGTGGACTGGTGC 52  
QY 61 GAGGACAACCTACACCATCGTGCCTGCTATCGCCGAGTTCTACACACGAT 110  
Db 51 GAGGACAACCTACACCATCGTGCCTGCTATCGCCGAGTTCTACACACGAT 2

RESULT 2  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F.G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
;
US-08-232-463-14
```

```
Query Match 6.0%; Score 49.4; DB 1; Length 7218;
Best Local Similarity 6.1%; Pred. No. 0.00016;
Matches 26; Conservative 218; Mismatches 179; Indels 0; Gaps 0;

QY 369 TAGGTTCAAGTGGTGGTCAAGTCTCTGCTGGGTTACGACGTGCTGCAATTGTGCAA 428
Db 1014 TAGGCTACAGAATTAAATCCGAGCTGGCTGAGTCCGAGGAGCTTGGATYYYYYY 1073

QY 429 GCTGCGCATCAACAACATCTCTGTGATGACCCCTGGAGTCTTGCACATGCTCAT 488
Db 1074 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1133

QY 489 CGCAGAGCTAAGAGGTGTGACACATGCTGTGTTAAGCTGGGCTCTTCTCGGCT 548
Db 1134 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1193

QY 549 CTGGTGGACCTGGCCCTGTTCTGCTGATCAGTACCGAGCTTCTCGAGCTGCTGC 608
Db 1194 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1253

QY 609 ATCCTTCAACTCCCTACCTGACATGCTGATGGCAATCCTCATCTGCTGCTGCTA 668
Db 1254 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1313

QY 669 CTGGGCTGTGATGCTTGCCTACTTGTATGCTGCTCAGATTCCTGAGCAAGCCC 728
Db 1314 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1373

QY 729 TGTCTCAAGTCTGGCCCAATGAGAAATGGGCTTCTATGTTGCCCTATGTCCT 788
Db 1374 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1433

QY 789 CCT 791
Db 1434 YGT 1436
```

```
RESULT 3
US-09-007-005-17/c
; Sequence 17, Application US/09007005B
; Patent No. 6258558
```

```
;
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
;
US-09-007-005-17

Query Match 4.9%; Score 40.8; DB 3; Length 289;
Best Local Similarity 6.9%; Pred. No. 0.011;
Matches 15; Conservative 95; Mismatches 106; Indels 0; Gaps 0;

QY 507 TGACACATGCGTGTGTTAAGCTGGGCTCTTCTCGGCTCTGCTGGACCTGGCCCT 566
Db 233 YGYCAAYGYCYTYGYSYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNY 174

QY 567 GTTCTCTCGATCAGTGACGAGCTTCTCGAGCTGCTGCATCTTCAACTTCCCTA 626
Db 173 YNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNY 114

QY 627 CTTGACTGATGCTGACATCTCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTT 686
Db 113 YNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYNYCYA 54

QY 687 TGCTACTTGTGCTGCTCCTCAGATTCCTGAGCA 722
Db 53 YTTGTGTATYATYTYGTYAYATTAAGTYA 18

RESULT 4
US-09-244-796-17/c
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
```

```
;
;
; FEATURE:
; OTHER INFORMATION: Translation template
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match      4.9%; Score 40.8; DB 3; Length 289;
Best Local Similarity 6.9%; Pred. No. 0.011;
Matches 15; Conservative 95; Mismatches 106; Indels 0; Gaps 0;

QY 507 TGACAAATGCGTGTGTTAAGCTGGGCTCTTCGCGGCTCTGGTGACCCCTGGCCCT 566
Db 233 YGYAYGCTGYGYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYN 174
QY 567 GTTCTGCTGGATCAGTGACCGAGCTTTCTGCGAGCTGCTGCATCCTTCACTTCCCTA 626
Db 173 YNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYN 114
QY 627 CTGCACTGATGGGACATCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 686
Db 113 YNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYNYSYN 54
QY 687 TGCTACTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 722
Db 53 YTYGYAYAYTYTYGYTYAYAYAYAYAYAYAYAYAYAYAYAYAYAYAYAYAYAY 18

RESULT 5
US-09-799-451-875
; Sequence 875, Application US/09799451
; Patent No. 6783969
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyang
; APPLICANT: Zhang, Jie
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ma, Yunqing
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Dunrui
; APPLICANT: Yang, Yonghong
; APPLICANT: Wehrman, Tom
; APPLICANT: Ghosh, Reena
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6783969el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/09/799,451
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 948
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 875
; LENGTH: 1063
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)...(802)
US-09-799-451-875

Query Match      4.9%; Score 40.2; DB 4; Length 1063;
Best Local Similarity 57.6%; Pred. No. 0.035;
Matches 72; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 7 GCCCGCACTGGTGGGACCACTGCAGCTGGTAGCTCGGAGGTGGACTGGTGGAGGAC 66
```

```
;
;
; FEATURE:
; OTHER INFORMATION: Translation template
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match      4.9%; Score 37.4; DB 3; Length 2514;
Best Local Similarity 50.9%; Pred. No. 0.41;
Matches 89; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 499 AAGAGGTGTGACAAACATGCGTGTGTTTAAAGCTGGGCTCTTCTCGGCGCTCTG 558
Db 570 ATGCGGCGCGCGAGGTGTCATGGCCAAACATGCTGCTCATCGGCTTCTCTCGTG 629
QY 559 CTGGCCCTGTTCTGCTGGATCAGTACCGAGCTTCTGCGAGCTGCTGCTCATCT 618
Db 630 AGCAGCTGTGCTGCTGCGCGCGCTTCTTCCCACTACGAGCACTGACCTTCTCC 689
QY 619 TTCCCTTACCTGCACTGCACTGCGGACATCTCATCTGCTGCTGCTGCTGCTG 673
Db 690 GCCTACTACTGCTTCTCATCCCTTACCACCATCGGCTTGGGAGTACGCTGG 744

RESULT 7
US-09-489-039A-106
; Sequence 106, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLBBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
```

```
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 106
; LENGTH: 1020
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-106

Query Match      4.3%; Score 35.8; DB 4; Length 1020;
Best Local Similarity 50.3%; Pred. No. 0.77;
Matches 88; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 486 CATCGCAGAGCTAAGACAGGTGTGACACATGCTGTGTAGCTGGGCTCTTCGCGG 545
Db 417 CCTCGCGCCCTGCAGGCGGAGACAGACGCTGGGGCTGATTCGTCCGGGATCATTC 476

QY 546 CCTCTGGTGGACCTCGCCCTGTCTGTCTGGATCAGTGACCGAGCTTCTGCGAGTCT 605
Db 477 CAGCGCTTCTTCGCGGCGCTGTGTAGCTGTAGCTGATGCGCTGCTGCCGACGAGAGAC 536

QY 606 GTCACTCTTCAACTTCCCTTACCTGCACTGCATGTGGCACATCTCTCATCTGCTT 660
Db 537 GTTACCGAATCGTTTCTGTGGCTGTCTCGCAGCTTCGGCAGCCGCGAGCTGGCAT 591
```

```
RESULT 8
US-09-475-515-67
; Sequence 67, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 67
; LENGTH: 2340
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp160.modUS4.delVI/V2
US-09-475-515-67

Query Match      4.3%; Score 35.4; DB 4; Length 2340;
Best Local Similarity 57.8%; Pred. No. 1.6;
Matches 63; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 12 GCACCTGGTGGACACGCTGCTGAGCTGGTAGCTCGGAGGTGGAGTGGTGGAGCAACTA 71
Db 2133 GTACTGGTGAACCTGCTGAGTACTGGAGCCAGGAGCTGAAGAGCAGCGCCGTGAGCCT 2192

QY 72 CACCATCGTGCCTATCGCGAGTCTTACAAACAGATCAGCAATGTC 120
Db 2193 GTTCAACGCCACCGCCATCGCCGTGGCGGAGGACCGCCGCGATCATC 2241
```

```
RESULT 9
US-09-475-515-68
; Sequence 68, Application US/09475515A
; Patent No. 6602705
```

```
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 68
; LENGTH: 2385
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp160.modUS4.del 128-194
US-09-475-515-68
```

```
Query Match      4.3%; Score 35.4; DB 4; Length 2385;
Best Local Similarity 57.8%; Pred. No. 1.6;
Matches 63; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 12 GCACCTGGTGGACACGCTGCTGAGCTGGTAGCTCGGAGGTGGAGTGGTGGAGCAACTA 71
Db 2178 GTACTGGTGAACCTGCTGAGTACTGGAGCCAGGAGCTGAAGAGCAGCGCCGTGAGCCT 2237

QY 72 CACCATCGTGCCTATCGCGAGTCTTACAAACAGATCAGCAATGTC 120
Db 2238 GTTCAACGCCACCGCCATCGCCGTGGCGGAGGACCGCCGCGATCATC 2286
```

```
RESULT 10
US-09-475-515-65
; Sequence 65, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475,515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 65
; LENGTH: 2538
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gp160.modUS4.delVI
US-09-475-515-65
```

```
Query Match      4.3%; Score 35.4; DB 4; Length 2538;
Best Local Similarity 57.8%; Pred. No. 1.7;
Matches 63; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
```



APPLICANT: HARTOG, Karin  
APPLICANT: LIU, Hong  
APPLICANT: GREER, Catherine  
APPLICANT: SELBY, Mark  
APPLICANT: WALKER, Christopher  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
FILE REFERENCE: 1621.002  
CURRENT APPLICATION NUMBER: US/09/475,515A  
CURRENT FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 73  
LENGTH: 4766  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
OTHER INFORMATION: gp160.modUSA.gag.modSF2  
US-09-475-515-73

Query Match 4.3%; Score 35.4; DB 4; Length 4766;  
Best Local Similarity 57.8%; Pred. No. 2.4;  
Matches 63; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 12 GCATGCTGGGACCCAGCTGCGAGCTGCTAGCTCGGAGGTGGACTGGTGGCGAGGACACTA 71  
Db 2427 GTACTGTGGAACCTGCTGAGTACTGAGCCGAGGAGCTGAAGAGAGCGCCGAGCCT 2486

QY 72 CACCATCGTCGCTATCGCCGAGTTCTACAAACAGCATCAGCAATGTC 120  
Db 2487 GTTCAACGCCACCGCCATCGCCGAGGCGGAGGACCGACCGCATCATC 2535

RESULT 15  
US-08-485-588-1  
Sequence 1, Application US/08485588  
Patent No. 5688938  
GENERAL INFORMATION:  
APPLICANT: Edward M. Brown  
APPLICANT: Steven C. Hebert  
APPLICANT: Forrest H. Fuller  
APPLICANT: James E. Garrett, Jr.  
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
TITLE OF INVENTION: MOLECULES  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: First Interstate World Center  
STREET: Suite 4700  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: FASTSEQ  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,588  
FILING DATE: 7 June, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below: 9  
APPLICATION NUMBER: 08/353,784  
FILING DATE: 9 December, 1994  
APPLICATION NUMBER: PCT/US/94/12117  
FILING DATE: 21 October, 1994  
APPLICATION NUMBER: U.S. 08/292,827  
FILING DATE: 23 August, 1994

APPLICATION NUMBER: U.S. 08/141,248  
FILING DATE: 22 October, 1993  
APPLICATION NUMBER: U.S. 08/009,389  
FILING DATE: 23 February, 1993  
APPLICATION NUMBER: U.S. 08/017,127  
FILING DATE: 12 February, 1993  
APPLICATION NUMBER: U.S. 07/934,161  
FILING DATE: 21 August, 1992  
APPLICATION NUMBER: U.S. 07/834,044  
FILING DATE: 11 February, 1992  
APPLICATION NUMBER: U.S. 07/749,451  
FILING DATE: 23 August, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Heber, Sheldon O.  
REGISTRATION NUMBER: 38,179  
REFERENCE/DOCKET NUMBER: 213/005  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5275 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 515..3769  
OTHER INFORMATION:  
US-08-485-588-1

Query Match 4.3%; Score 35.2; DB 1; Length 5275;  
Best Local Similarity 50.6%; Pred. No. 3;  
Matches 85; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 483 GCTCATCGCAGAGCTAAAGAGGTGTGACAAACATGCGTGTGTTAAGCTGGGCCCTTCTC 542  
Db 2437 GCCATCGTCAAGCCACCAACCGGAGCTCTCCTATCTCTTCTTCTTCTTCTGCTG 2496

QY 543 GGGCCTTGTGTGACCCCTGGCCCTGTTCTGTCGATCAGTGACCGAGCTTTTCTCGAGCT 602  
Db 2497 CTGCTTCTCCAGCTCCCTGTTCTTCATCGGGGAGCCCGAGCTGGAGCTGCGCCCTGCG 2556

QY 603 GCTGTCTATCTTCAACTTCCCTACCTGCACTGCACTGTCGATCTCT 650  
Db 2557 CCAGCCGGCCCTTTGGCATCAGCTTCGTGCTCTGATCTCTGTCATCT 2604

Search completed: November 21, 2004, 07:04:51  
Job time : 75.8343 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 20:53:12 ; Search time 386.775 Seconds  
(without alignments)  
11567.962 Million cell updates/sec

Title: US-10-017-410-3

Perfect score: 828

Sequence: 1 atggcgcccgccactggtg.....catcagtcacagatcacgtga 828

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3627888 seqs, 2701811610 residues

Total number of hits satisfying chosen parameters: 7255776

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 19: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 20: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 21: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description          |
|------------|-------|-------------|--------|----|----------------------|
| 1          | 828   | 100.0       | 828    | 13 | US-10-017-410-3      |
| 2          | 813.6 | 98.3        | 4202   | 16 | US-10-188-832-21     |
| 3          | 813.6 | 98.3        | 4212   | 15 | US-10-295-027-151    |
| 4          | 761   | 91.9        | 822    | 15 | US-10-182-447-6      |
| 5          | 664.8 | 80.3        | 4175   | 13 | US-10-017-410-1      |
| 6          | 350.6 | 42.3        | 744    | 10 | US-09-945-527-48     |
| 7          | 183.4 | 22.1        | 792    | 15 | US-10-182-447-4      |
| 8          | 52.2  | 6.3         | 35425  | 15 | US-10-017-161-2429   |
| 9          | 52.2  | 6.3         | 35425  | 15 | US-10-292-798-2069   |
| 10         | 42.2  | 5.1         | 3034   | 18 | US-10-425-115-77895  |
| 11         | 41.8  | 5.0         | 1604   | 17 | US-10-437-963-77989  |
| 12         | 41    | 5.0         | 637    | 18 | US-10-425-115-146588 |

|   |    |      |     |        |    |                     |                   |
|---|----|------|-----|--------|----|---------------------|-------------------|
| c | 13 | 40.2 | 4.9 | 284    | 15 | US-10-002-631C-17   | Sequence 17, Appl |
|   | 14 | 40.2 | 4.9 | 636    | 16 | US-10-264-237-615   | Sequence 615, App |
|   | 15 | 40.2 | 4.9 | 801    | 15 | US-10-182-447-5     | Sequence 5, Appli |
|   | 16 | 40.2 | 4.9 | 1063   | 16 | US-10-302-172-875   | Sequence 875, App |
|   | 17 | 39.2 | 4.7 | 598    | 17 | US-10-767-701-20737 | Sequence 20737, A |
|   | 18 | 39   | 4.7 | 1194   | 16 | US-10-425-114-32295 | Sequence 32295, A |
|   | 19 | 39   | 4.7 | 1429   | 18 | US-10-425-115-77897 | Sequence 77897, A |
|   | 20 | 39   | 4.7 | 1471   | 17 | US-10-767-701-12261 | Sequence 12261, A |
|   | 21 | 39   | 4.7 | 1547   | 17 | US-10-029-386-27071 | Sequence 27071, A |
| c | 22 | 38.8 | 4.7 | 332    | 15 | US-10-029-386-27071 | Sequence 27071, A |
|   | 23 | 38.8 | 4.7 | 579    | 15 | US-10-029-386-27071 | Sequence 27071, A |
|   | 24 | 38.8 | 4.7 | 744802 | 15 | US-10-252-798-13371 | Sequence 13371, A |
| c | 25 | 38.6 | 4.7 | 357    | 18 | US-10-425-115-24309 | Sequence 24309, A |
|   | 26 | 38   | 4.6 | 158405 | 14 | US-10-175-523-86    | Sequence 86, Appl |
|   | 27 | 37.4 | 4.5 | 854    | 9  | US-09-746-491-55    | Sequence 55, Appl |
|   | 28 | 37.4 | 4.5 | 2514   | 9  | US-09-939-484-3     | Sequence 3, Appli |
|   | 29 | 37.4 | 4.5 | 2514   | 9  | US-09-939-484-3     | Sequence 3, Appli |
|   | 30 | 37.4 | 4.5 | 2590   | 15 | US-10-146-733-7     | Sequence 7, Appli |
| c | 31 | 37.2 | 4.5 | 700    | 16 | US-10-260-238-3722  | Sequence 3722, Ap |
|   | 32 | 37.2 | 4.5 | 25576  | 10 | US-09-764-891-8306  | Sequence 8306, Ap |
| c | 33 | 37   | 4.5 | 851    | 18 | US-10-425-115-53388 | Sequence 53388, A |
| c | 34 | 36.8 | 4.4 | 3657   | 16 | US-10-388-934-165   | Sequence 165, App |
| c | 35 | 36.8 | 4.4 | 3657   | 16 | US-10-388-934-165   | Sequence 165, App |
|   | 36 | 36.6 | 4.4 | 607    | 14 | US-10-123-155-344   | Sequence 344, App |
|   | 37 | 36.6 | 4.4 | 607    | 15 | US-10-146-731-344   | Sequence 344, App |
|   | 38 | 36.6 | 4.4 | 607    | 15 | US-10-140-472-344   | Sequence 344, App |
|   | 39 | 36.6 | 4.4 | 607    | 15 | US-10-141-761-344   | Sequence 344, App |
|   | 40 | 36.6 | 4.4 | 607    | 15 | US-10-142-885-344   | Sequence 344, App |
|   | 41 | 36.6 | 4.4 | 607    | 15 | US-10-158-790-344   | Sequence 344, App |
|   | 42 | 36.6 | 4.4 | 607    | 15 | US-10-137-871-344   | Sequence 344, App |
|   | 43 | 36.6 | 4.4 | 607    | 15 | US-10-140-923-344   | Sequence 344, App |
|   | 44 | 36.6 | 4.4 | 607    | 15 | US-10-141-756-344   | Sequence 344, App |
|   | 45 | 36.6 | 4.4 | 607    | 15 | US-10-141-759-344   | Sequence 344, App |

#### ALIGNMENTS

RESULT 1  
US-10-017-410-3  
; Sequence 3, Application US/10017410  
; Publication No. US20020115094A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnham, Peggy J  
; TITLE OF INVENTION: Polynucleotide Differentially Expressed in Liver Cancer  
; FILE REFERENCE: 960296.97401  
; CURRENT APPLICATION NUMBER: US/10/017,410  
; CURRENT FILING DATE: 2001-12-14  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 828  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(825)  
US-10-017-410-3

Query Match 100.0%; Score 828; DB 13; Length 828;  
Best Local Similarity 100.0%; Pred. No. 2.2e-253;  
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|    |     |   |     |
|----|-----|---|-----|
| QY | 1   | ATGGCGCCCGCCACTGTTGGGACCCAGCTCGAGCTGTAGCTCGAGGTGGAGTGTGTC   | 60  |
| Db | 1   | ATGGCGCCCGCCACTGTTGGGACCCAGCTCGAGCTGTAGCTCGAGGTGGAGTGTGTC   | 60  |
| QY | 61  | GAGGACAACATACACCATCGTCTGCTATCGCGAGTTCTACACACGATCAGCAATGTC   | 120 |
| Db | 61  | GAGGACAACATACACCATCGTCTGCTATCGCGAGTTCTACACACGATCAGCAATGTC   | 120 |
| QY | 121 | TTATTTTTCATTATACGCCCATCTGCTGCTGTTTGTAGTATGATGATGCAACATGCTTG | 180 |

121 TTATTTTCAATTTACCGCCCATCTGATGCTTGTGATGAGTATCAACATGCTTG 180  
181 AACAGTGACATCTACTTAATCTGAGCTCTTTTGGTTGATGGGAATGGATCCGCTAC 240  
181 AACAGTGACATCTACTTAATCTGAGCTCTTTTGGTTGATGGGAATGGATCCGCTAC 240  
241 TTCCATTTTACCCCTTAGTTTCTTTGGGTGAGATGCTTGTGATGAATTCAGTCCCTTTGGGTT 300  
241 TTCCATTTTACCCCTTAGTTTCTTTGGGTGAGATGCTTGTGATGAATTCAGTCCCTTTGGGTT 300  
301 CTGATGTGCTTTGGCCATGCTGTTCCCGAGAGTATCTACCAAGATCTTTCCGAAT 360  
301 CTGATGTGCTTTGGCCATGCTGTTCCCGAGAGTATCTACCAAGATCTTTCCGAAT 360  
361 GACAGGGGTAGGTTCAAGGTGGTGGTCAAGTGTCTCTGCGGTACGAGTGCCTGGCA 420  
361 GACAGGGGTAGGTTCAAGGTGGTGGTCAAGTGTCTCTGCGGTACGAGTGCCTGGCA 420  
421 TTGTCAAGCCTGCCATCAACAACTCTCTCTGATGACCCCTGGAGTTCCTTGCACTGCA 480  
421 TTGTCAAGCCTGCCATCAACAACTCTCTCTGATGACCCCTGGAGTTCCTTGCACTGCA 480  
481 CTGCTCATCGCAGAGCTTAAGAGGTGTGCAACATGCTGTGTTTAAAGCTGGGCTCTTC 540  
481 CTGCTCATCGCAGAGCTTAAGAGGTGTGCAACATGCTGTGTTTAAAGCTGGGCTCTTC 540  
541 TCGGGCTCTGCTGGGACCTCGGCTCTGCTGATCAGTGAACCGAGCTTTCTGCGAG 600  
541 TCGGGCTCTGCTGGGACCTCGGCTCTGCTGATCAGTGAACCGAGCTTTCTGCGAG 600  
601 CTGCTGTCACTCTTCAACTTCCCTTACCTGCACTGCAATGAGGACATCTCTGCTGCT 660  
601 CTGCTGTCACTCTTCAACTTCCCTTACCTGCACTGCAATGAGGACATCTCTGCTGCT 660  
661 GCTGCTTACCTGGCTGTGATGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 720  
661 GCTGCTTACCTGGCTGTGATGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 720  
721 CAAGGCTCTGCTCAATCAAGTTCTGCGCCCAATGAGAAATGGGCTTCAATGCTGCCCTAT 780  
721 CAAGGCTCTGCTCAATCAAGTTCTGCGCCCAATGAGAAATGGGCTTCAATGCTGCCCTAT 780  
781 GTGTCCCTCTGCTGTGCAACAGAAATCATCAAGTCAAGATCAAGTCA 828  
781 GTGTCCCTCTGCTGTGCAACAGAAATCATCAAGTCAAGATCAAGTCA 828

RESULT 2  
US-10-188-832-21  
; Sequence 21, Application US/10188832  
; Publication No. US20040076955A1  
; GENERAL INFORMATION:  
; APPLICANT: Mack, David H.  
; APPLICANT: Aziz, Natasha  
; TITLE OF INVENTION: Methods of Diagnosis of Bladder Cancer, Compositions  
; TITLE OF INVENTION: and Methods of Screening for Modulators of Bladder  
; TITLE OF INVENTION: Cancer  
; FILE REFERENCE: 018501-002330US  
; CURRENT APPLICATION NUMBER: US/10/188,832  
; CURRENT FILING DATE: 2002-11-22  
; PRIOR APPLICATION NUMBER: US 60/302,814  
; PRIOR FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: US 60/310,099  
; PRIOR FILING DATE: 2001-08-03  
; PRIOR APPLICATION NUMBER: US 60/343,705  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/350,666  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/372,246  
; PRIOR FILING DATE: 2002-04-12  
; NUMBER OF SEQ ID NOS: 207

; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 21  
; LENGTH: 4202  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-188-832-21  
  
Query Match 98.3%; Score 813.6; DB 16; Length 4202;  
Best Local Similarity 98.9%; Pred. No. 2e-246; Indels 0; Gaps 0;  
Matches 819; Conservative 0; Mismatches 9;  
  
QY 1 ATGGGGCGCCCGCAGCTGGTGGGACAGCTGCGAGCTGGTAGCTCGAGAGTGGACTGGTGC 60  
Db 77 ATGGGGCGCCCGCAGCTGGTGGGACAGCTGCGAGCTGGTAGCTCGAGAGTGGACTGGTGC 136  
QY 61 GAGGACAACTACACCATCGTGGCTGTATCGCGAGTTCACAAACAGATCAGCAATGTC 120  
Db 137 GAGGACAACTACACCATCGTGGCTGTATCGCGAGTTCACAAACAGATCAGCAATGTC 196  
QY 121 TTATTTTTCATTTTACGGCCCATCTGATGCTGCTGTTGATGAGTATGCAACATGCTTG 180  
Db 197 TTATTTTTCATTTTACGGCCCATCTGATGCTGCTGTTGATGAGTATGCAACATGCTTG 256  
QY 181 AACAGTGACATCTACTTAATCTGGAATCTTTTGGTTGTAGTGGGAATTTGATCCGCTAC 240  
Db 257 AACAGTGACATCTACTTAATCTGGAATCTTTTGGTTGTAGTGGGAATTTGATCCGCTAC 316  
QY 241 TTCCATTTTACCCCTTAGTTTCTTTGGGTGAGATGCTTGAATTCAGTCCCTTTGGGTT 300  
Db 317 TTCCATTTTACCCCTTAGTTTCTTTGGGTGAGATGCTTGAATTCAGTCCCTTTGGGTT 376  
QY 301 CTGATGTGCTTTGGCCATGCTGTTCCCGAGAGTATCTACCAAGATCTTTTCGGAAT 360  
Db 377 CTGATGTGCTTTGGCCATGCTGTTCCCGAGAGTATCTACCAAGATCTTTTCGGAAT 436  
QY 361 GACAGGGGTAGGTTCAAGGTGGTGGTCAAGTGTCTCTGCTGCGGTACGAGTGCCTGGCA 420  
Db 437 GACAGGGGTAGGTTCAAGGTGGTGGTCAAGTGTCTCTGCTGCGGTACGAGTGCCTGGCA 496  
QY 421 TTGTCAAGCCTGCCATCAACAACTCTCTGATGACCCCTGGAGTTCCTTGCACTGCA 480  
Db 497 TTGTCAAGCCTGCCATCAACAACTCTCTGATGACCCCTGGAGTTCCTTGCACTGCA 556  
QY 481 CTGCTCATCGCAGAGCTAAAGAGGTGTGCAACATGCTGTGTTAAAGCTGGGCTCTTC 540  
Db 557 CTGCTCATCGCAGAGCTAAAGAGGTGTGCAACATGCTGTGTTAAAGCTGGGCTCTTC 616  
QY 541 TCGGGCTCTGCTGGGACCCCTGGCCCTGTTCTGCTGGATCAGTGACCGAGCTTTCTGGAG 600  
Db 617 TCGGGCTCTGCTGGGACCCCTGGCCCTGTTCTGCTGGATCAGTGACCGAGCTTTCTGGAG 676  
QY 601 CTGCTGTCACTCTTCAACTTCCCTTACCTGCACTGCAATGAGGACATCTCTCATCTGCT 660  
Db 677 CTGCTGTCACTCTTCAACTTCCCTTACCTGCACTGCAATGAGGACATCTCTCATCTGCT 736  
QY 661 GCTGCTTACCTGGCTGTGATGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 720  
Db 737 GCTGCTTACCTGGCTGTGATGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 796  
QY 721 CAAGGCTCTGCTCAATCAAGTTCTGCGCCCAATGAGAAATGGGCTTCAATGCTGCCCTAT 780  
Db 797 CAAGGCTCTGCTCAATCAAGTTCTGCGCCCAATGAGAAATGGGCTTCAATGCTGCCCTAT 856  
QY 781 GTGTCCCTCTGCTGTGCAACAGAAATCATCAAGTCAAGATCAAGTCA 828  
Db 857 GTGTCCCTCTGCTGTGCAACAGAAATCATCAAGTCAAGATCAAGTCA 904

RESULT 3  
US-10-295-027-151  
; Sequence 151, Application US/10295027  
; Publication No. US2003023250A1  
; GENERAL INFORMATION:

APPLICANT: Afar, Daniel  
APPLICANT: Aziz, Natasha  
APPLICANT: Ginsberg, Wendy M.  
APPLICANT: Gish, Kurt C.  
APPLICANT: Glynn, Richard  
APPLICANT: Hevezi, Peter A.  
APPLICANT: Mack, David H.  
APPLICANT: Murray, Richard  
APPLICANT: Watson, Susan R.  
APPLICANT: Eos Biotechnology, Inc.  
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
TITLE OF INVENTION: Methods of Screening for Modulators of Cancer  
FILE REFERENCE: 018501-012500US  
CURRENT APPLICATION NUMBER: US/10/295,027  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: US 09/663,733  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: US 60/350,666  
PRIOR FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 60/335,394  
PRIOR FILING DATE: 2001-11-15  
PRIOR APPLICATION NUMBER: US 60/332,464  
PRIOR FILING DATE: 2001-11-21  
PRIOR APPLICATION NUMBER: US 60/334,393  
PRIOR FILING DATE: 2001-11-29  
PRIOR APPLICATION NUMBER: US 60/340,376  
PRIOR FILING DATE: 2001-12-14  
PRIOR APPLICATION NUMBER: US 60/347,211  
PRIOR FILING DATE: 2002-01-08  
PRIOR APPLICATION NUMBER: US 60/347,349  
PRIOR FILING DATE: 2002-01-10  
PRIOR APPLICATION NUMBER: US 60/355,250  
PRIOR FILING DATE: 2002-02-08  
PRIOR APPLICATION NUMBER: US 60/356,714  
PRIOR FILING DATE: 2002-02-13  
Remaining Prior Application data removed - See File wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1386  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 151  
LENGTH: 4212  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-295-027-151

Query Match 98.3%; Score 813.6; DB 15; Length 4212;  
Best Local Similarity 98.9%; Pred. No. 2e-248;  
Matches 819; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
Qy 1 ATGGGCGCCCGCAGCTGGTGGGACCACTGCAGGCTGTAGCTCGGAGGTGGACTGGTGC 60  
Db 87 ATGGGCGCCCGCAGCTGGTGGGACCACTGCAGGCTGTAGCTCGGAGGTGGACTGGTGC 146  
Qy 61 GAGGACAATACCACTCGTGCCTGTCTACCGAGTCTTACAAACGATCAGCAATGTC 120  
Db 147 GAGGACAATACCACTCGTGCCTGTCTACCGAGTCTTACAAACGATCAGCAATGTC 206  
Qy 121 TTATTTTTCATTTACCGCCCACTGCATGTGCTTTGATGAGTATGCAATGCTTG 180  
Db 207 TTATTTTTCATTTACCGCCCACTGCATGTGCTTTGATGAGTATGCAATGCTTG 266  
Qy 181 AACAGTCACATCTACTTAATCTGACCTTTTGGTGTAGTGGAAATGGATCGGTCTAC 240  
Db 267 AACAGTCACATCTACTTAATCTGACCTTTTGGTGTAGTGGAAATGGATCGGTCTAC 326  
Qy 241 TTCCATTTTACCTTTAGTTTCTTGGGTGATGCTTGAATCTTGCAGTCTTTTGGGTT 300  
Db 327 TTCCATGCAACCTTAGTTTCTTGGGTGATGCTTGAATCTTGCAGTCTTTTGGGTT 386  
Qy 301 CTGATGTGCTTTGGCCATGTGTTCCCGAAGGATCTTACCAAGATCTTTTCGGAAT 360  
Db 387 CTGATGTGCTTTGGCCATGTGTTCCCGAAGGATCTTACCAAGATCTTTTCGGAAT 446  
Qy 361 GACAGGGGTAGTTCAAGGTGGTGGTGCAGTGTCTCTGCTCGGTTACGAGTGCCTGGCA 420

Db 447 GACCGGGTAGTTCAAGGTGGTGGTCTCAGTGTCTCTGCTGGGTACGACGTGCTGGCA 506  
Qy 421 TTTGTCAAGCTGCGCATCAACAAATCTCTCTGATGATCCCTGGGAGTTCCTTTGCACTGCA 480  
Db 507 TTTGTCAAGCTGCGCATCAACAAATCTCTCTGATGATCCCTGGGAGTTCCTTTGCACTGCA 566  
Qy 481 CTGCTCATGCGCAGAGCTAAAGAGGTGTGACAAATGCGTGTGTTTAAAGCTGGGCTCTTC 540  
Db 567 CTGCTCATGCGCAGAGCTAAAGAGGTGTGACAAATGCGTGTGTTTAAAGCTGGGCTCTTC 626  
Qy 541 TCGGGCTCTGGTGGACCTGGCCCTGTTCTGCTGATGATGATGATGATGATGATGATGATG 600  
Db 627 TCGGGCTCTGGTGGACCTGGCCCTGTTCTGCTGATGATGATGATGATGATGATGATGATG 686  
Qy 601 CTGCTGTCTCTCTCAACTTCCCTTACCTGACATGATGATGATGATGATGATGATGATGATG 660  
Db 687 CTGCTGTCTCTCTCAACTTCCCTTACCTGACATGATGATGATGATGATGATGATGATGATG 746  
Qy 661 GCTGCTTACCTGGCTGTGATGCTTTTGCCTTACTTTGATGCTGCTCAGAGATTCCTGAG 720  
Db 747 GCTGCTTACCTGGCTGTGATGCTTTTGCCTTACTTTGATGCTGCTCAGAGATTCCTGAG 806  
Qy 721 CAAGGCTCTCTCATCAAGTTCGCGCAATGAGAAATGGGCTTCAATGCTGCTGCTGCTGCT 780  
Db 807 CAAGGCTCTCTCATCAAGTTCGCGCAATGAGAAATGGGCTTCAATGCTGCTGCTGCTGCT 866  
Qy 781 GTGTCTCTCTCTGCTGCGCAACAAAGAAATCATCAAGTCAAGTCAAGTCAAGTCAAGTCA 828  
Db 867 GTGTCTCTCTCTGCTGCGCAACAAAGAAATCATCAAGTCAAGTCAAGTCAAGTCAAGTCA 914

RESULT 4  
US-10-182-447-6  
; Sequence 6, Application US/10182447  
; Publication No. US20030185814A1  
; GENERAL INFORMATION:  
; APPLICANT: RADT, Marcus  
; TITLE OF INVENTION: CERAMIDASE  
; FILE REFERENCE: P68055USO  
; CURRENT APPLICATION NUMBER: US/10/182,447  
; CURRENT FILING DATE: 2002-07-29  
; PRIOR APPLICATION NUMBER: PCT/EP01/00900  
; PRIOR FILING DATE: 2001-01-27  
; PRIOR APPLICATION NUMBER: DE 10003293.1  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: DE 10011392.3  
; PRIOR FILING DATE: 2000-03-09  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 822  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-182-447-6

Query Match 91.9%; Score 761; DB 15; Length 822;  
Best Local Similarity 96.0%; Pred. No. 5.4e-232;  
Matches 792; Conservative 0; Mismatches 30; Indels 3; Gaps 1;  
Qy 1 ATGGGCGCCCGCAGCTGGTGGGACCACTGCAGGCTGTAGCTCGGAGGTGGACTGGTGC 60  
Db 1 ATGGGCGCCCGCAGCTGGTGGGACCACTGCAGGCTGTAGCTCGGAGGTGGACTGGTGC 60  
Qy 61 GAGGACAATACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
Db 61 GAGGACAATACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
Qy 121 TTATTTTTCATTTTACCGCCCACTGCATGTGCTTTGATGAGTATGCAATGCTTG 180  
Db 121 TTATTTTTCATTTTACCGCCCACTGCATGTGCTTTGATGAGTATGCAATGCTTG 180





```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21462)..(21603)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23918)..(24055)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (26460)..(26597)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (26460)..(26597)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (26868)..(27016)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (35125)..(35225)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (35270)..(35425)
; OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-2429

Query Match          6.3%; Score 52.2; DB 15; Length 35425;
Best Local Similarity 61.3%; Pred. No. 7.9e-05;
Matches 84; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 232 TCCGCTCTACTTCCATTTTACCCCTTAGTTCTTGGGTCAGATGCTTGATGAATTGCGAGTC 291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21470 TCCATGTATTTCCACATGACGCTCAGCTTCTCTGGCCAGCTGCTGGACGAGATCGCCATC 21529

QY 292 CTTTGGGTTCTGATGTGTGCTTTTGGCCATGTGTTCCCAAGAAGGTATCTACCAAGATC 351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21530 CTGTGGCTCTGGGCGAGTGGCTATAGCATATGATGCGCCGCTGCTATTTTCCCTCCTTC 21589

QY 352 TTTCGGAATGACAGGGG 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21590 CTTGGGGGGAACAGGTG 21606

RESULT 9
US-10-292-798-2069
; Sequence 2069, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: SUWA, MAKIKO
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ASURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2069
; LENGTH: 35425
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(35425)
; NAME/KEY: CDS
; LOCATION: (201)..(293)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21253)..(21367)
; FEATURE:
; NAME/KEY: CDS
```

```
; LOCATION: (21462)..(21603)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (23918)..(24055)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (26460)..(26597)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (26868)..(27016)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (35125)..(35225)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (35270)..(35369)
; OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-2069

Query Match          6.3%; Score 52.2; DB 15; Length 35425;
Best Local Similarity 61.3%; Pred. No. 7.9e-05;
Matches 84; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 232 TCCGCTCTACTTCCATTTTACCCCTTAGTTCTTGGGTCAGATGCTTGATGAATTGCGAGTC 291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21470 TCCATGTATTTCCACATGACGCTCAGCTTCTCTGGCCAGCTGCTGGACGAGATCGCCATC 21529

QY 292 CTTTGGGTTCTGATGTGTGCTTTTGGCCATGTGTTCCCAAGAAGGTATCTACCAAGATC 351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21530 CTGTGGCTCTGGGCGAGTGGCTATAGCATATGATGCGCCGCTGCTATTTTCCCTCCTTC 21589

QY 352 TTTCGGAATGACAGGGG 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21590 CTTGGGGGGAACAGGTG 21606

RESULT 10
US-10-425-115-77895
; Sequence 77895, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 77895
; LENGTH: 3034
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_171063C.1
US-10-425-115-77895

Query Match          5.1%; Score 42.2; DB 18; Length 3034;
Best Local Similarity 61.3%; Pred. No. 0.036;
Matches 68; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 33 GGCTGGTAGCTCGGAGTGGACTGTGGAGGACAACTACACCATGTGGCTGCTATCGC 92
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 GCCTGTACATCAACGATTGAGTTGTGGAGGAGAATTATGCACATCATATATATAGC 460

QY 93 CGAGTTCTACACACGATCAGCAATGTCTTATTTTCAATTTTACCGCCCAT 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 AGAATTCTACACACAAATTTCTAATGCCATGTGTTCTTTTGGCGCTTAT 511

RESULT 11
```

```
US-10-437-963-77989/c
; Sequence 77989, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Brad
; APPLICANT: Li, Bing
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 77989
; LENGTH: 1604
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1604)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_77836C.1
US-10-437-963-77989

Query Match          5.0%; Score 41.8; DB 17; Length 1604;
Best Local Similarity 52.6%; Pred. No. 0.035;
Matches 91; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 513 CATCGGTGTTAAAGCTGGGCTCTTCGGGCTCTGGGACCTGGGCTGTTCTG 572
Db 1375 CATACGGTGTGAGGCTCATCTCTCTCAGATCGGTCGAGTCCGGCGGATCAG 1316
Qy 573 CTGATCAGTGACCGAGCTTTCGAGCTGCTGTATCTTCATCTTCAACTTCCCTACCTGCA 632
Db 1315 CCGCATCTTCGGCGCGCTTCGCGAGCTCTGACGACGTCAACGTCAGCCGACCGAC 1256
Qy 633 CTGCATGTGGACATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 685
Db 1255 CTCCTTCATCCCTCCGACGTGAAGCTGCCTCCCAACCTTGGATCCGGTCACT 1203

RESULT 12
US-10-425-115-146588
; Sequence 146588, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 146588
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_65204C.1
US-10-425-115-146588

Query Match          5.0%; Score 41; DB 18; Length 637;
Best Local Similarity 52.7%; Pred. No. 0.039;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

US-10-437-963-77989/c
; Sequence 77989, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Brad
; APPLICANT: Li, Bing
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 77989
; LENGTH: 1604
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1604)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_77836C.1
US-10-437-963-77989

Query Match          5.0%; Score 41.8; DB 17; Length 1604;
Best Local Similarity 52.6%; Pred. No. 0.035;
Matches 91; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 513 CATCGGTGTTAAAGCTGGGCTCTTCGGGCTCTGGGACCTGGGCTGTTCTG 572
Db 1375 CATACGGTGTGAGGCTCATCTCTCTCAGATCGGTCGAGTCCGGCGGATCAG 1316
Qy 573 CTGATCAGTGACCGAGCTTTCGAGCTGCTGTATCTTCATCTTCAACTTCCCTACCTGCA 632
Db 1315 CCGCATCTTCGGCGCGCTTCGCGAGCTCTGACGACGTCAACGTCAGCCGACCGAC 1256
Qy 633 CTGCATGTGGACATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 685
Db 1255 CTCCTTCATCCCTCCGACGTGAAGCTGCCTCCCAACCTTGGATCCGGTCACT 1203

RESULT 12
US-10-425-115-146588
; Sequence 146588, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 146588
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_65204C.1
US-10-425-115-146588

Query Match          5.0%; Score 41; DB 18; Length 637;
Best Local Similarity 52.7%; Pred. No. 0.039;
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

US-10-002-631C-17/c
; Sequence 17, Application US/10002631C
; Publication No. US20030157486A1
; GENERAL INFORMATION:
; APPLICANT: Graff, Jonathon M.
; APPLICANT: Muenster, Matthew
; TITLE OF INVENTION: METHODS TO IDENTIFY SIGNAL SEQUENCES
; FILE REFERENCE: A34943 090495.0243
; CURRENT APPLICATION NUMBER: US/10/002,631C
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/300,309
; PRIOR FILING DATE: 2001-06-21
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 284
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-002-631C-17

Query Match          4.9%; Score 40.2; DB 15; Length 284;
Best Local Similarity 57.6%; Pred. No. 0.047;
Matches 72; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 7 GCCCGCCAGCTGTTGGGACCGAGCTGAGCTGAGTGGAGTGGAGTGGAGTGGAGGAC 66
Db 222 GCCCGGACCGAGAGGGCTACTGGGGCCCGACGACCTCCACGCTGGAGCTGGTGGAGGAG 163
Qy 67 AACTACACCATCGTGGCTGCTATCGCGAGTCTTACACAGATCAGCAATGCTTATTT 126
Db 162 AACTACTCCGTGACCTGCTATCATCGCGAGTCTTGAATACAGTGAGTAACCTGATCATG 103
Qy 127 TTCTAT 131
Db 102 ATTAT 98

RESULT 14
US-10-264-237-615
; Sequence 615, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131p1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 615
; LENGTH: 636
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
Qy 527 AGCTGGGCCCTCTTCTCGGGCTCTGTTGGAGCCCTGGCCCTGTTCTGCTGGATCATGACC 586
Db 88 AGCGCGCGCGAGGCGGCTCGCGGGGACCAACGCGACGTCCTCAAGCTCACACACC 147
Qy 587 GAGCTTTCTGCGAGCTGCTGCATCTTCAACTTCCCTTACCTGCTGCTGCTGCTGCTGCTG 646
Db 148 GCGCTCTCTCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 207
Qy 647 TCCTCATCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 695
Db 208 GCTTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 256

RESULT 13
US-10-002-631C-17/c
; Sequence 17, Application US/10002631C
; Publication No. US20030157486A1
; GENERAL INFORMATION:
; APPLICANT: Graff, Jonathon M.
; APPLICANT: Muenster, Matthew
; TITLE OF INVENTION: METHODS TO IDENTIFY SIGNAL SEQUENCES
; FILE REFERENCE: A34943 090495.0243
; CURRENT APPLICATION NUMBER: US/10/002,631C
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/300,309
; PRIOR FILING DATE: 2001-06-21
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 284
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-002-631C-17

Query Match          4.9%; Score 40.2; DB 15; Length 284;
Best Local Similarity 57.6%; Pred. No. 0.047;
Matches 72; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 7 GCCCGCCAGCTGTTGGGACCGAGCTGAGCTGAGTGGAGTGGAGTGGAGTGGAGGAC 66
Db 222 GCCCGGACCGAGAGGGCTACTGGGGCCCGACGACCTCCACGCTGGAGCTGGTGGAGGAG 163
Qy 67 AACTACACCATCGTGGCTGCTATCGCGAGTCTTACACAGATCAGCAATGCTTATTT 126
Db 162 AACTACTCCGTGACCTGCTATCATCGCGAGTCTTGAATACAGTGAGTAACCTGATCATG 103
Qy 127 TTCTAT 131
Db 102 ATTAT 98

RESULT 14
US-10-264-237-615
; Sequence 615, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131p1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 615
; LENGTH: 636
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

Job time : 389.775 secs

```
; NAME/KEY: misc feature
; LOCATION: (560)..(560)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (605)..(605)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (631)..(631)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-237-615
```

```
Query Match 4.9%; Score 40.2; DB 16; Length 636;
Best Local Similarity 57.6%; Pred. No. 0.07;
Matches 72; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 7 GCCCGGACCTGGTGGGACACGAGCTGCAGGCTGGTAGCTCGGAGGTGGACTGGTGGAGGAC 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 GCCGGGACCGAGAGGGCTACTGGGGCCCGACGACCTCCACGCTGGACTGGTGGAGGAG 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 67 AACTACACCATCGTGCCTGCTATCGCGGAGTTCTTACAACAGATCAGCAATGTCATTAT 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 AACTACTCCGTGACCTGGTACATCGCGGAGTTCTTGAATACAGTGAGTAACCTGATCATG 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 127 TTCAT 131
   ||| |||
Db 180 ATTAT 184
```

```
RESULT 15
US-10-182-447-5
; Sequence 5, Application US/10182447
; Publication No. US20030185814A1
; GENERAL INFORMATION:
; APPLICANT: RADT, Marcus
; APPLICANT: HOFMANN, Kay
; TITLE OF INVENTION: CERAMIDASE
; FILE REFERENCE: P68055US0
; CURRENT APPLICATION NUMBER: US/10/182,447
; CURRENT FILING DATE: 2002-07-29
; PRIOR APPLICATION NUMBER: PCT/EP01/00900
; PRIOR FILING DATE: 2001-01-27
; PRIOR APPLICATION NUMBER: DE 10003293.1
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: DE 10011392.3
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-182-447-5
```

```
Query Match 4.9%; Score 40.2; DB 15; Length 801;
Best Local Similarity 57.6%; Pred. No. 0.079;
Matches 72; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 7 GCCCGGACCTGGTGGGACACGAGCTGCAGGCTGGTAGCTCGGAGGTGGACTGGTGGAGGAC 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 GCCGGGACCGAGAGGGCTACTGGGGCCCGACGACCTCCACGCTGGACTGGTGGAGGAG 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 67 AACTACACCATCGTGCCTGCTATCGCGGAGTTCTTACAACAGATCAGCAATGTCATTAT 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 AACTACTCCGTGACCTGGTACATCGCGGAGTTCTTGAATACAGTGAGTAACCTGATCATG 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 127 TTCAT 131
   ||| |||
Db 130 ATTAT 134
```

Search completed: November 21, 2004, 07:44:02



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 20:50:47 ; Search time 2284.57 Seconds  
(without alignments)  
13206.895 Million cell updates/sec

Title: US-10-017-410-3

Perfect score: 828  
Sequence: 1 atgggcgcgcgcactggtg.....catcagtcacagatcacgtga 828

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gss1.\*  
9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID      | Description        |
|------------|-------|-------------|--------|------------|--------------------|
| 1          | 671.4 | 81.1        | 681    | 9 AY401889 | AY401889 Homo sapi |
| 2          | 660.2 | 79.7        | 681    | 9 AY401890 | AY401890 Pan trogl |
| 3          | 593.6 | 71.7        | 1022   | 5 BU511164 | BU511164 AGNCOURT  |
| 4          | 535.4 | 64.7        | 681    | 9 AY401891 | AY401891 Mus muscu |
| 5          | 512.4 | 61.9        | 868    | 6 CA976684 | CA976684 AGNCOURT  |
| 6          | 512.4 | 61.9        | 1173   | 3 AK085306 | AK085306 Mus muscu |
| 7          | 480   | 58.0        | 797    | 6 CA463294 | CA463294 AGNCOURT  |
| 8          | 454.4 | 54.9        | 565    | 6 CF115220 | CF115220 Shultzomi |
| 9          | 447.4 | 54.0        | 449    | 5 BX646596 | BX646596 DKFZp781B |
| 10         | 427.2 | 51.6        | 633    | 2 BB660847 | BB660847           |
| 11         | 416.2 | 50.3        | 498    | 4 BT48265  | BT48265 470659 MA  |
| 12         | 414.8 | 50.1        | 1527   | 3 AF370405 | AF370405 Homo sapi |
| 13         | 403   | 48.7        | 698    | 5 BU234223 | BU234223 603792086 |
| 14         | 381   | 46.0        | 781    | 5 BU232554 | BU232554 603409105 |
| 15         | 372.6 | 45.0        | 548    | 6 CF169369 | CF169369 B0812G07- |
| 16         | 361.4 | 43.6        | 847    | 5 BU220481 | BU220481 603107516 |
| 17         | 348.8 | 42.1        | 678    | 5 BU203269 | BU203269 604153603 |
| 18         | 343.6 | 41.5        | 937    | 5 BX842808 | BX842808 BX842808  |
| 19         | 341.4 | 41.2        | 432    | 6 CF169808 | CF169808 B081BD08- |
| 20         | 340.6 | 41.1        | 648    | 4 BJ060795 | BJ060795 BJ060795  |
| 21         | 329.4 | 39.8        | 454    | 6 CB784796 | CB784796 AMGNNUC:T |
| 22         | 325.2 | 39.3        | 427    | 1 AA900336 | AA900336 UI-R-EQ-C |
| 23         | 324.2 | 39.2        | 401    | 2 BF554219 | BF554219 UI-R-CO-h |
| 24         | 306.4 | 37.0        | 434    | 6 CB758986 | CB758986 AMGNNUC:M |

#### ALIGNMENTS

RESULT 1  
AY401889  
LOCUS Homo sapiens HCM1041 gene, VIRTUAL TRANSCRIPT, partial sequence, 681 bp DNA linear GSS 12-DEC-2003  
DEFINITION Homo sapiens HCM1041 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.  
ACCESSION AY401889  
VERSION AY401889.1 GI:39757875  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 681)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 681)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.  
FEATURES  
source Location/Qualifiers  
1..681  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
<1..>681  
/locus\_tag="HCM1041"

Query Match 81.1%; Score 671.4; DB 9; Length 681;

Best Local Similarity 99.1%; Pred. No. 2.7e-180;

Matches 675; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 148 ATGTGCTTGTGATGATGACATGCTTGACAGTGCATCTACTTAATCTGACT 207

Db 1 ATGTGCTTGTGCTGATGACATGCTTGACAGTGCATCTACTTAATCTGACT 60

QY 208 CTTTGGTTGTTAGTGGGAATGGATCGCTACTTCCATTTTACCTTAGTTCTTGGGT 267  
Db 61 CTTTGGTTGTTAGTGGGAATGGATCGCTACTTCCATTTTACCTTAGTTCTTGGGT 120  
QY 268 CAGATCGTTGATCAACTTTCAGTCCCTTTGGTCTGATGTGCTTTGGCCATGTGGTTC 327  
Db 121 CAGATCGTTGATCAACTTTCAGTCCCTTTGGTCTGATGTGCTTTGGCCATGTGGTTC 180  
QY 328 CCAGAAGGTATCTACCAAGATCTTTTCGGAATGACAGGGGTAGGTTCAAGGTGGTGGTC 387  
Db 181 CCAGAAGGTATCTACCAAGATCTTTTCGGAATGACAGGGGTAGGTTCAAGGTGGTGGTC 240  
QY 388 AGTGTCTGTCTGCGGTTCAGAGTGTGCTGGCAATTTGTTCAGGCTGCGCATCAACAATC 447  
Db 241 AGTGTCTGTCTGCGGTTCAGAGTGTGCTGGCAATTTGTTCAGGCTGCGCATCAACAATC 300  
QY 448 TCTCTGATGACCTCGGAGTTCTTTCGACATGCACTGCTCATCGCAGAGCTAAAGAGGTGT 507  
Db 301 TCTCTGATGACCTCGGAGTTCTTTCGACATGCACTGCTCATCGCAGAGCTAAAGAGGTGT 360  
QY 508 GACAAATGCGTGTGTTTAAAGTCTGCGCTCTTCTCGGGCTCTGTTGGACCTTGGCCCTG 567  
Db 361 GACAAATGCGTGTGTTTAAAGTCTGCGCTCTTCTCGGGCTCTGTTGGACCTTGGCCCTG 420  
QY 568 TTCTGTGATGACCTGACCGAGCTTTCTGAGTGTGCTGCTCATCGCAGAGCTAAAGAGGTGT 627  
Db 421 TTCTGTGATGACCTGACCGAGCTTTCTGAGTGTGCTGCTCATCGCAGAGCTAAAGAGGTGT 480  
QY 628 CTGCACTGATGCGGACATCTCTCATCTGCTTGTCTGCTTACCTGCGCTGTGATGCTTT 687  
Db 481 CTGCACTGATGCGGACATCTCTCATCTGCTTGTCTGCTTACCTGCGCTGTGATGCTTT 540  
QY 688 GCTTACTTTGATGCTGCTCAGAGATCTTCGACAAAGGCGCTGCTCATCAAGTTCTGCCCC 747  
Db 541 GCTTACTTTGATGCTGCTCAGAGATCTTCGACAAAGGCGCTGCTCATCAAGTTCTGCCCC 600  
QY 748 AATGAGAAATGGGCTTCATGTTGTTGCTTCCCTATGTTGCTTCTGTCGCAACAGAAA 807  
Db 601 AATGAGAAATGGGCTTCATGTTGTTGCTTCCCTATGTTGCTTCTGTCGCAACAGAAA 660  
QY 808 TCATCAGTCAAGATCAGTGA 828  
Db 661 TCATCAGTCAAGATCAGTGA 681

RESULT 2  
AY401890  
LOCUS  
DEFINITION Pan troglodytes HCM1041 gene, VIRTUAL TRANSCRIPT, partial sequence,  
genomic survey sequence.  
ACCESSION  
VERSION AY401890.1 GI:39757876  
KEYWORDS  
SOURCE GSS.  
ORGANISM Pan troglodytes (chimpanzee)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
REFERENCE  
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,  
Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,  
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 681)  
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,  
Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,  
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Direct Submission

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source 1..681  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
<1..>681  
/locus\_tag="HCM1041"  
gene  
ORIGIN  
Query Match 79.7%; Score 660.2; DB 9; Length 681;  
Best Local Similarity 98.1%; Pred. No. 4.2e-177;  
Matches 668; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 148 ATGTGCTTGTGATGAGTATGCAACATGCTTGAACAGTGAACATCTACTTAATCTGGACT 207  
Db 1 ATGTGCTTGTGATGAGTATGCAACATGCTTGAACAGTGAACATCTACTTAATCTGGACT 60  
QY 208 CTTTGGTTGTTAGTGGGAATTTGGATCCGTCTACTTCCATTTTACCCCTTAGTTTCTTGGGT 267  
Db 61 CTTTGGTTGTTAGTGGGAATTTGGATCCGTCTACTTCCATGCAACCCCTTAGTTTCTTGGGT 120  
QY 268 CAGATCGTTGATGAACTTTCGAGTCTTTCGGTCTGATGTGCTTTGGCCATGTGGTTC 327  
Db 121 CAGATCGTTGATGAACTTTCGAGTCTTTCGGTCTGATGTGCTTTGGCCATGTGGTTC 180  
QY 328 CCAGAAGGTATCTACCAAGATCTTTTCGGAATGACAGGGGTAGGTTCAAGGTGGTGGTC 387  
Db 181 CCAGAAGGTATCTACCAAGATCTTTTCGGAATGACAGGGGTAGGTTCAAGGTGGTGGTC 240  
QY 388 AGTGTCTGTCTGCGGTTCAGAGTGTGCTGGCAATTTGTTCAGGCTGCGCATCAACAATC 447  
Db 241 AGTGTCTGTCTGCGGTTCAGAGTGTGCTGGCAATTTGTTCAGGCTGCGCATCAACAATC 300  
QY 448 TCTCTGATGACCTGCGGAGTTCTTTCGACATGCACTGCTCATCGCAGAGCTAAAGAGGTGT 507  
Db 301 TCTCTGATGACCTGCGGAGTTCTTTCGACATGCACTGCTCATCGCAGAGCTAAAGAGGTGT 360  
QY 508 GACAAATGCGTGTGTTTAAAGTCTGCGCTCTTCTCGGGCTCTGTTGGACCTTGGCCCTG 567  
Db 361 GACAAATGCGTGTGTTTAAAGTCTGCGCTCTTCTCGGGCTCTGTTGGACCTTGGCCCTG 420  
QY 568 TTCTGTGATGACCTGCGGAGTTCTTTCGAGTGTGCTGCTCATCAAGTTCTGCCCC 627  
Db 421 TTCTGTGATGACCTGCGGAGTTCTTTCGAGTGTGCTGCTCATCAAGTTCTGCCCC 480  
QY 628 CTGCACTGATGCGGACATCTCTCATCTGCTTGTCTGCTTACCTGCGCTGTGATGCTTT 687  
Db 481 CTGCACTGATGCGGACATCTCTCATCTGCTTGTCTGCTTACCTGCGCTGTGATGCTTT 540  
QY 688 GCTTACTTTGATGCTGCTCAGAGATCTTCGACAAAGGCGCTGCTCATCAAGTTCTGCCCC 747  
Db 541 GCTTACTTTGATGCTGCTCAGAGATCTTCGACAAAGGCGCTGCTCATCAAGTTCTGCCCC 600  
QY 748 AATGAGAAATGGGCTTCATGTTGTTGCTTCCCTATGTTGCTTCTGTCGCAACAGAAA 807  
Db 601 AATGAGAAATGGGCTTCATGTTGTTGCTTCCCTATGTTGCTTCTGTCGCAACAGAAA 660  
QY 808 TCATCAGTCAAGATCAGTGA 828  
Db 661 TCATCAGTCAAGATCAGTGA 681

RESULT 3  
BUS11164  
LOCUS  
DEFINITION AGENCOURT\_10107530 NIH\_MGC\_134 Mus musculus cdna clone  
IMAGE:6505924 5', mRNA sequence.  
ACCESSION BUS11164  
VERSION BUS11164.1 GI:22817397





|           |  |
|-----------|--|
| ORGANISM  | Mus musculus   |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.   |
| AUTHORS   | 1 Carninci, P. and Hayashizaki, Y.   |
| TITLE     | High-efficiency full-length cDNA cloning   |
| JOURNAL   | Meth. Enzymol. 303, 19-44 (1999)   |
| MEDLINE   | 99279253   |
| PUBMED    | 10349636   |
| REFERENCE | 2  |
| AUTHORS   | Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.   |
| TITLE     | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes   |
| JOURNAL   | Genome Res. 10 (10), 1617-1630 (2000)  |
| MEDLINE   | 20499374   |
| PUBMED    | 11042159   |
| REFERENCE | 3  |
| AUTHORS   | Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kusunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  |
| TITLE     | RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer   |
| JOURNAL   | Genome Res. 10 (11), 1757-1771 (2000)  |
| MEDLINE   | 20530913   |
| PUBMED    | 11076861   |
| REFERENCE | 4  |
| AUTHORS   | The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.   |
| TITLE     | Functional annotation of a full-length mouse cDNA collection   |
| JOURNAL   | Nature 409, 685-690 (2001)   |
| REFERENCE | 5  |
| AUTHORS   | The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.   |
| TITLE     | Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs   |
| JOURNAL   | Nature 420, 563-573 (2002)   |
| REFERENCE | 6 (bases 1 to 1173)  |
| AUTHORS   | Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. |
| TITLE     | Direct Submission  |
| JOURNAL   | Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  |
| COMMENT   | cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/  |
| FEATURES  | Location/Qualifiers  |
| source    | 1..1173 /organism="Mus musculus"   |

|          |   |
|----------|---|
|          | /mol_type="mRNA"  |
|          | /strain="C57BL/6J"  |
|          | /db_xref="FANTOM DB:D630008P07"   |
|          | /db_xref="taxon:10090"  |
|          | /clone="D630008P07"   |
|          | /tissue_type="kidney"   |
|          | /clone_lib="RIKEN full-length enriched mouse cDNA library"  |
|          | /dev_stage="0 day neonate"  |
|          | 72..731   |
| CDS      | /note="unnamed protein product; putative similar to CANCER RELATED GENE-LIVER 1 [Mus musculus] (SPTR AAL40408, evidence: FASTY, 100%ID, 77.8%length, match=642)"  |
|          | /codon_start=1  |
|          | /protein_id="BAC39416.1"  |
|          | /db_xref="GI:26351559"  |
|          | /translation="MGAPHWDLRAGSSEVDWEDNYTIVPAIEFYNTINSLVFFI LPPICMCLPROYATCFNSGIVLITLLVVVGIGSVYPHATISFLGMDLDELAILVLM CALAMWFRERYLPKIFRNDGRFKAIVCVLSAITTCLAFIRPAINNISIMILGLECTA LLVIELKRCNDNRVRFKGLGSLWTLALFCWISDOAFCELLSFFHPYLHCWSADR G" |
| ORIGIN   |   |
|          | Query Match 61.9%; Score 512.4; DB 3; Length 1173;  |
|          | Best Local Similarity 87.4%; Pred. No. 8.4e-135;  |
|          | Matches 561; Conservative 0; Mismatches 81; Indels 0; Gaps 0;   |
| QY       | 1 ATGGGGCGGGCCGCACTGGTGGGACCCAGCTGCGAGGCTGGTAGCTCGGAGGTGACCTGGTGC 60  |
| DB       |   |
| QY       | 72 ATGGGGCGGGCCGCACTGGTGGGACCCAGCTGCGAGGCTGGTAGCTCGGAGGTGACCTGGTGC 131  |
| DB       |   |
| QY       | 61 GAGGCAACTACACCATCGTCTCTATCGCCGAGTTCTACAACACGATCAGCAATGTC 120   |
| DB       |   |
| QY       | 132 GAGGCAACTACACCATCGTCTCTATCGCCGAGTTCTACAACACGATCAGCAATGTC 191  |
| DB       |   |
| QY       | 121 TTATTTTTCATTTTACCGCCCATCTGCATGTCCTGTTTGTATGAGTATCATCAATGCTTC 180  |
| DB       |   |
| QY       | 192 TTGTTTTCATTTTACCTCCCATCTGCATGTCCTGTTTCCCGCAGTACCAACGCTGCTTC 251   |
| DB       |   |
| QY       | 181 AACAGTGCATCTACTTAATCTGGAATCTTTTGGTTGTAGTGGGAATTCGATCCGCTAC 240  |
| DB       |   |
| QY       | 252 AACAGGCGCATCTACTTAATCTGGAATCTTTTGGTTGTAGTGGGAATTCGATCTGCTAC 311   |
| DB       |   |
| QY       | 241 TTCCATTTTACCCCTAGTTTCTTGGGTGAGTGCATGCTGATGAACTTCCTTGGGTT 300  |
| DB       |   |
| QY       | 312 TTCCATGCAACGCTGAGTTTCTCGGTGAGTGCATGCTGATGAATCTGCTGCTTCTGCTG 371   |
| DB       |   |
| QY       | 301 CTGATGCTGCTTTGGCCATCTGTTTCCCGAAGGATCTACCAAGATCTTCGGAAT 360  |
| DB       |   |
| QY       | 372 CTGATGCTGCTTTGGCCATCTGTTTCCCGAAGGATCTACCAAGATCTTCGGAAT 431  |
| DB       |   |
| QY       | 361 GACAGGGGTAGTCTCAAGTGGTGGTTCAGTGCCTGCTGCGGTTACGACGCTGCTGGCA 420  |
| DB       |   |
| QY       | 432 GACAGGGGTAGTCTCAAGTGGTGGTTCAGTGCCTGCTGCGGTTACGACGCTGCTGGCG 491  |
| DB       |   |
| QY       | 421 TTGTGCAAGCTTGCATCAACCAATCTCTCTGATGACCCCTGGGAGTTCCTTGCATGCA 480  |
| DB       |   |
| QY       | 492 TTATCAAGCCCGCCATCAACCAATTTTCCCTGATGATTTCTGGGACTTCCATGCACTGG 551   |
| DB       |   |
| QY       | 481 CTGCTCATCGGAGCTTAAGAGGTTGTCACACATGCGTGTGTTTAAGCTGGGCTCTTC 540   |
| DB       |   |
| QY       | 552 CTGCTTGTGCAAGCTGAAGAGGTTGTCACCAATGTCGCTGTGTTTAAAGCTGGGCTCTTC 611  |
| DB       |   |
| QY       | 541 TCGGGCTCTGCTGGTGCACCTGGCCCTGTTCTGCTGCATCAGTACCGAGCTTCTTGCAG 600   |
| DB       |   |
| QY       | 612 TCTGGCTCTGCTGGTGCATCTGGCTCTTCTGCTGCATCAGGACCAAGCCCTTCTGTGAG 671   |
| DB       |   |
| QY       | 601 CTGCTGTGATCTTCAACTTCCCTACCTGCACTGCATGTGG 642  |
| DB       |   |
| QY       | 672 CTGCTCTCTCTCTTCACTTCCCTACCTGCACTGTGTGG 713  |
| DB       |   |
| RESULT 7 |   |

CA463294 797 bp mRNA linear EST 12-NOV-2002  
 LOCUS AGNCOURT 10691314 NIH\_MGC\_169 Mus musculus cDNA clone  
 IMAGE:6770649 5', mRNA sequence.  
 ACCESSION CA463294  
 VERSION CA463294.1 GI:24919646  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 797)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Dr. Jonathan Kuo, NIMH  
 cDNA Library Preparation: Michael Brownstein Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1CM3088 row: p column: 08  
 High quality sequence stop: 536.  
 FEATURES  
 source Location/Qualifiers  
 1..797  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clones="IMAGE:6770649"  
 /lab\_host="DH10B (TI-phage-resistant)"  
 /clone\_lib="NIH\_MGC\_169"  
 /note="Organ: Testicles; Vector: pDNR-LIB; Site 1: SfII  
 (ggccattagggcc); Site 2: SfiI (ggccgctcgcc); cDNA made  
 by oligo-dT priming and directionally cloned. 5' and 3'  
 adaptors were used in cloning as follows:  
 5'-AAGCAGTGGTATCAACGAGATGGCCATTACGCCGGG-3' and  
 5'-ATTCTAGAGCGGAGCGGCGACATG-dt(30)NN-3'. Full-length  
 enriched library was constructed using the Clontech  
 Creator SMART kit and size-selected to contain the 0.5 kb  
 size fraction. Library created in the laboratory of M.  
 Brownstein (NIMH, NIH). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 58.0%; Score 480; DB 6; Length 797;  
 Best Local Similarity 87.0%; Pred. No. 1.3e-125;  
 Matches 550; Conservative 0; Mismatches 80; Indels 2; Gaps 2;  
 QY 1 ATGGGCGCCCGCACTGGTGGACCACTGCTAGCTGAGGTGGAGTGGTGC 60  
 DB 35 ATGGGCGCCCGCACTGGTGGACCACTGCTAGCTGAGGTGGAGTGGTGC 94  
 QY 61 GAGGACAACTACCACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 120  
 DB 95 GAGGACAACTACCACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 154  
 QY 121 TTATTTTTCATTTTACCGCCCATCTGATGTGCTTGTGATGATGATGATGATG 180  
 DB 155 TTGTTTTCATTTTACCTCCCATCTGACGTGTTGTTCCGCCAGTACGCAAGTGTTC 214  
 QY 181 AACAGTGACATCTACTTAATCTGGACTCTTTTGGTGTGTAGTGGAAATGGATCGTCTAC 240  
 DB 215 AACAGCGGATCTACTTAATATGACCGCTCTAGTTGTAGTGGGATGGATCTGTCTAC 274  
 QY 241 TTCATTTTACCTTAGTTTCTGGGTGAGATGCTTGTGATGAATTCGAGTCTTTGGGTT 300  
 DB 275 TTCCATGCAACGCTGAGTTTCTGGGTGAGATGCTTGTGATGAATTCGAGTCTTTGGGTT 334  
 QY 301 CTGATGTGTCTTTGGCCATCTGTTCCCGAGAGTATCTACCAAGATCTTTCCGAAT 360  
 DB 335 CTGATGTGTCTTTGGCCATCTGTTTCCCGAGAGGATTTTACCAAGATCTTTCCGAAT 394

QY 361 GACAGGGGTAGTTTCAAGGTGGTGTGCTAGTCTCTGTCTGCGGTTACGACGTGCCTGGCA 420  
 DB 395 GACAGGGGCAGGTTCAAGGGCAGTGGTGTGCTGTCTGCAATTTACAACTGCTGGCG 454  
 QY 421 TTTTGTCAAGCCTGCCATCAACAAATCTCTGTATGACCTCGGAGTTCCTTTGCACTGCA 480  
 DB 455 TTTATCAAGCCCGCATCAACAATATTTCCCTGATGATTTCTGGGACTTCATGCACTGG 514  
 QY 481 CTGCTCATCGCAGAGCTAAAGAGGTGTGACAACTGCGTGTGTTTAACTGGGCTCTTTC 540  
 DB 515 CTGCTTGTGTGAGAGCTCAAGAGGTGTGACAACTGCGTGTGTTTAACTGGGCTCTTTC 574  
 QY 541 TCGGGCCTCTGTGTGAGCCCTGGCCCTGTCTGCTGATCACTGACCGAG-CTTTTCTGGA 599  
 DB 575 TCTGGCCTCTGTGTGAGCTCTGGCTCTCTTCTGTGGATCAGCAGCAAGCCCTTCTGTGA 634  
 QY 600 GCTGCTGT-CATCCTTCAACTTCCCTACCTG 630  
 DB 635 GCTGCTCTCCCTCTTTCACCTTCCCTACCTG 666

RESULT 8  
 CF115220 565 bp mRNA linear EST 23-JUL-2003  
 LOCUS Shultzomica08471 Rat lung airway and parenchyma cDNA libraries  
 DEFINITION Rattus norvegicus cDNA clone NP6159 5', mRNA sequence.  
 ACCESSION CF115220  
 VERSION CF115220.1 GI:33175919  
 KEYWORDS EST.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 565)  
 AUTHORS Shultz, M.A., Zhang, L., Gu, Y.-Z., Baker, G.L., Fannuchi, M.V.,  
 Padua, A.M., Gurske-W.A., Morin, D., Penn, S.G., Jovanovich, S.B.,  
 Plopper, C.G. and Buckpitt, A.R.  
 TITLE Gene expression analysis in response to lung toxicants: I.  
 Sequencing and microarray development  
 JOURNAL Am. J. Respir. Cell Mol. Biol. 30 (3), 296-310 (2004)  
 COMMENT Contact: Shultz MA  
 Dept. of Molecular Biosciences, School of Veterinary Medicine  
 University of California, Davis  
 1311 Haring Hall, One Shields Avenue, Davis, CA 95616, USA  
 Tel: 530 752 0793  
 Fax: 530 752 4698  
 Email: [mashultz@ucdavis.edu](mailto:mashultz@ucdavis.edu)  
 Average Phred score is 20 or better. All poor quality data (Phred <  
 20) and vector/linker sequence has been removed.  
 High quality sequence stop: 565.

## FEATURES

source Location/Qualifiers  
 1..565  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="NP6159"  
 /sex="male"  
 /tissue type="airway or parenchyma"  
 /dev stage="adult"  
 /clone\_lib="Rat lung airway and parenchyma cDNA libraries"  
 /note="Organ: lung; Vector: pGEM-11zf(-); Site 1: Eco RI;  
 Site 2: Not I; mRNA was isolated from microdissected rat  
 lung airways and parenchyma tissues."

## ORIGIN

Query Match 54.9%; Score 454.4; DB 6; Length 565;  
 Best Local Similarity 88.8%; Pred. No. 2.5e-118;  
 Matches 502; Conservative 0; Mismatches 62; Indels 1; Gaps 1;  
 QY 218 TAGTGGGAATGGATCGTCTACTTCCATT-TTACCCTTAGTTTCTTTGGGTGAGATGCTT 276

```
|||||
Db 1 TAGTGGGAATCGGATCTGTCTACTTCCATNGCAACTCTTAGTTCTCCCTGGGTCAGATGCTT 60
QY 277 GATCAACTTCGAGTCCCTTTGGTGTCTGATGTGCTTTGGCCATGTGCTTCCCGAGAGG 336
Db 61 GATGAATCTGCAATCTTTGGTGTCTGATGTGCTTTGGCCATGTGCTTCCCGAGAGG 120
QY 337 TATCTACCAAGATCTTTTGGGAATGACAGGGGTAGTTTCAAGGTGGTGGTTCAGTGTCTTG 396
Db 121 TATTTACCAAGATCTTTTGGGAATGACAGGGGTAGTTTCAAGGTGGTGGTTCAGTGTCTTG 180
QY 397 TCTCGGTTAGCAGTGGTGGCATTCTTCAAGCTGCCATCAACATCTCTCTGATG 456
Db 181 TCTCGAGTTACCAATGCTGGCGTTTGTCAAGCGCTGCTATCAACAATATCTCTCTGATG 240
QY 457 ACCCTGGGAGTTCCTTGGACATGCACTGCTCATCGCAGAGCTAAAGAGGTGTGACAAATG 516
Db 241 ATTCTGGGGGTTCCGTGACATGCGGTGCTCATTCGAGAGCTGAAGAGGTGTGACAAATG 300
QY 517 CGTGTGTTTAAAGCTGGGCTCTTCTCGGGCCCTCTGCTGAGCCCTGGCCCTGTTCTGCTGG 576
Db 301 CGTGTGTTTAAAGCTGGGCTCTTCTCTGCGCTTTTGGTGGACTCTGGCTCTCTCTGCTGG 360
QY 577 ATCAGTGACCGAGCTTCTCGAGTGTCTGTGATCTCTTCAACTTCCCTACCTGCACTGC 636
Db 361 ATCAGTGACCGAGCTTCTGTGAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
QY 637 ATGTGGACATCTCTCATCTGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 696
Db 421 GTGTGGACATCTCTCATCTGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
QY 697 GATGCTGCTCAGAGATCTCTGAGCAAGCCCTGTTCATCAAGTTCTGCGCCCAATGAGAAA 756
Db 481 GATGCTGCTCAGAGATCTCTGAGCAAGCCCTGTTCATCAAGTTCTGCGCCCAATGAGAAA 540
QY 757 TGGGCTTCATGTGTCCTCTATG 781
Db 541 TGGGCTTTTATCGGTGTCCCTATG 565
```

```
RESULT 9
BX646596 449 bp mRNA linear EST 04-SEP-2003
DEFINITION DKFZp781B0790_r1 781 (synonym: hlcc4) Homo sapiens cDNA clone
VERSION BX646596
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
```

```
REFERENCE 1 (bases 1 to 449)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
Fobo,G., Han,M. and Wiemann,S.
JOURNAL EST (Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., et al.)
COMMENT Unpublished (2003)
Contact: MIPS
```

```
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No SI sequence available.
This clone (DKFZp781B0790) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
```

```
FEATURES
source
1..449
/organism="Homo sapiens"
/mol_type="mRNA"
```

```
/db_xref="taxon:9606"
/clone="DKFZp781B0790"
/lab_host="adult"
/clone_lib="DH10B"
/clone_lib="781 (synonym: hlcc4)"
/note="Vector: pSport1_Sfi; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"
```

## ORIGIN

```
Query Match 54.0%; Score 447.4; DB 5; Length 449;
Best Local Similarity 99.8%; Pred. No. 2.3e-116;
Matches 448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 265 GGTGAGATGCTGTGATGAACCTTGGAGTCTCTTTGGGTTCTGATGTGCTTTGGCCATGCG 324
Db 1 GGTGAGATGCTGTGATGAACCTTGGAGTCTCTTTGGGTTCTGATGTGCTTTGGCCATGCG 60
QY 325 TTCCCGAGAGGTATCTACCAAGATCTTTCGGAATGACAGGGGTAGGTTCGAAGTGGTG 384
Db 61 TTCCCGAGAGGTATCTACCAAGATCTTTCGGAATGACAGGGGTAGGTTCGAAGTGGTG 120
QY 385 GTCAAGTGTCTGTCTGCGGTTTACGACGTGCTTGTCAAGCTTGTCAAGCTTGTCAAGCTTGT 444
Db 121 GTCAAGTGTCTGTCTGCGGTTTACGACGTGCTTGTCAAGCTTGTCAAGCTTGTCAAGCTTGT 180
QY 445 ATCTCTGTGATGACCTCGGAGTTCCTTGGACGTGCTGCTCATCGAGAGCTAAAGAGG 504
Db 181 ATCTCTGTGATGACCTCGGAGTTCCTTGGACGTGCTGCTCATCGAGAGCTAAAGAGG 240
QY 505 TGTGACACATGCGGTGCTTTAAAGCTGGGCTCTTCTCGGGCTCTGTTGGACCTGGCC 564
Db 241 TGTGACACATGCGGTGCTTTAAAGCTGGGCTCTTCTCGGGCTCTGTTGGACCTGGCC 300
QY 565 CTGTTCTGCTGGATCAGTGACCGAGCTTCTTCTGCGAGCTGCTGCTCATCTTCAACTTCCCC 624
Db 301 CTGTTCTGCTGGATCAGTGACCGAGCTTCTTCTGCGAGCTGCTGCTCATCTTCAACTTCCCC 360
QY 625 TACCTGCACTGATGTCACATCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 684
Db 361 TACCTGCACTGATGTCACATCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 685 TTTGCTACTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 713
Db 421 TTTGCTACTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
```

## RESULT 10

```
BX660847 633 bp mRNA linear EST 26-OCT-2001
DEFINITION BB660847 RIKEN full-length enriched, 0 day neonate kidney Mus
VERSION BB660847
KEYWORDS BB660847.1 GI:16494626
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 633)
AUTHORS Arakawa,T., Garninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
```

## TITLE

## JOURNAL

## COMMENT





/note=Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;  
Library made from pooled tissue from testis, thymus,  
semintendous muscle, longissimus muscle, pancreas,  
adrenal, and endometrium."

ORIGIN  
Query Match 50.3%; Score 416.2; DB 4; Length 498;  
Best Local Similarity 91.1%; Pred. No. 2e-107;  
Matches 453; Conservative 0; Mismatches 43; Indels 1; Gaps 1;  
QY 2 TGGGCGCCCGCAGCTGGTGGGACAGCTGAGGCTGGTAGCTCGGAGGTGGACTGGTGG 61  
Db 1 TGAGCGCCCGCAGCTGGTGGGACAGCTGCGGGTGGCAGCTCGGAGGTGGACTGGTGG 60  
QY 62 AGGACAACTACACCATGGTGGGCTGCTATCGCGAGTCTCAACACCATGAGCAATGCT 121  
Db 61 AGGACAACTACACCATGGTGGGCTGCTATCGCGAGTCTCAACACCATGAGCAATGCT 120  
QY 122 TATTTTTCATTTTACCGCCCATCTGCGATGCTGCTTGTGATGAGTATGCAACATGCTGA 181  
Db 121 TGTTTTTCATCTTGGCGCCCATCTGCGATGCTGCTTGTGATGAGTATGCAACATGCTGA 180  
QY 182 ACAGTGACATCTAATCTGGAATCTTTTGGTGTGATGGGAATGGATCGGCTACT 241  
Db 181 ACAGTGACATCTAATCTGGAATCTTTTGGTGTGATGGGAATGGATCGGCTACT 240  
QY 242 TCCATTTTACCTTTAGTTTCTTGGTGCAGATGCTTGTGATGAGTATGCAACATGCTGA 301  
Db 241 TCCATTTTACCTTTAGTTTCTTGGTGCAGATGCTTGTGATGAGTATGCAACATGCTGA 300  
QY 302 TGATGTGCTTGGTCCCATGCTTGGTCCCGAGAGTATCTACCAAGATCTTTCGGAATG 361  
Db 301 TGATGTGCTTGGTCCCATGCTTGGTCCCGAGAGTATCTACCAAGATCTTTCGGAATG 360  
QY 362 ACAGGGGTAGTTCAGGTGGTGGTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 421  
Db 361 ACAGGGGTAGTTCAGGTGGTGGTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420  
QY 422 TTGTCAAGCTGCTTACACATCTCTCATGACC-CTGGGAGTCTTCTGCACTGCA 480  
Db 421 TTGTCAAGCTGCTTACACATCTCTCATGACCCTGGGGGGGTTCCCTGCAATGGG 480  
QY 481 CTGCTCATCGCAGAGCT 497  
Db 481 CTACTCATGCTGAGCT 497

RESULT 12  
AF370405 1527 bp mRNA linear HTC 31-DEC-2003  
LOCUS Homo sapiens PP11646 mRNA, complete cds.  
DEFINITION AF370405  
ACCESSION AF370405.1 GI:33341739  
VERSION AF370405.1  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1527)  
AUTHORS Huang, Y., Zhou, X.M., Zhang, P.P., Jiang, H.Q., Qin, W.X., Zhao, X.T.,  
Wan, D.F. and Gu, J.R.  
Novel human cDNA clones with function of inhibiting cancer cell  
growth  
TITLE Unpublished  
JOURNAL 2 (bases 1 to 1527)  
REFERENCE Zhang, P.P., Zhou, X.M., Jiang, H.Q., Huang, Y., Qin, W.X., Zhao, X.T.,  
AUTHORS Wan, D.F. and Gu, J.R.  
Direct Submission  
TITLE Submitted (16-APR-2001) National Laboratory for Oncogenes & Related  
JOURNAL Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie-Tu Road, Shanghai  
200032, P. R. China  
FEATURES Location/Qualifiers  
1. .1527  
source

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
218. .640  
/codon\_start=1  
/product="PP11646"  
/protein\_id="AA015241.1"  
/db\_xref="GI:33341740"  
/translation="MCLFRYATCFNSGIYLIWLLVVGIGSVYPHATLSFLGMDL  
ELAVLVMCALAMWPPRRYLPKIFRNDGRKFRKVVSLSAVTTCCLAFVPAINNISL  
MTLGVPTALLIAELKHERNRRHRKGGQQGGGDKV"

CDS

ORIGIN

Query Match 50.1%; Score 414.8; DB 3; Length 1527;  
Best Local Similarity 82.4%; Pred. No. 7e-107;  
Matches 532; Conservative 0; Mismatches 17; Indels 97; Gaps 1;  
QY 108 GATCAGCAATGCTTATTTTTCATTTTACCGCCCATCTGCAATGCTGCTTGTGATGAGTGA 167  
Db 178 GATCAGCAATGCTTATTTTTCATTTTACCGCCCATCTGCAATGCTGCTTGTGATGAGTGA 237  
QY 168 TGCACATGCTTGAACAGTGCATCTACTTAATCTGCACTCTTTTGTGATGAGTGAAT 227  
Db 238 TGCACATGCTTGAACAGTGCATCTACTTAATCTGCACTCTTTTGTGATGAGTGAAT 297  
QY 228 TGGATCGCTCTACTTCCATTTTACCGCTTAGTTTCTTGGTGCAGATGCTTGTGATGAGTGA 287  
Db 298 TGGATCGCTCTACTTCCATTTTACCGCTTAGTTTCTTGGTGCAGATGCTTGTGATGAGTGA 357  
QY 288 AGTCTTTGGGTTCTGATGCTGCTTGGCCATGCTGCTTGGCCATGCTTGGCCATGCTTGGCC 347  
Db 358 AGTCTTTGGGTTCTGATGCTGCTTGGCCATGCTGCTTGGCCATGCTTGGCCATGCTTGGCC 417  
QY 348 GATCTTTCCGAATGACAGGGTAGGTTCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 407  
Db 418 GATCTTTCCGAATGACAGGGTAGGTTCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 477  
QY 408 GACGTGCTTGGCATTTGTCAAGCTGCTGCTCAACCAACATCTCTGATGACCTGGGAGT 467  
Db 478 GACGTGCTTGGCATTTGTCAAGCTGCTGCTCAACCAACATCTCTGATGACCTGGGAGT 537  
QY 468 TCCTTGGCATGCTGCTGCTCATCGCAGAGCTAAG----- 501  
Db 538 TCCTTGGCATGCTGCTGCTCATCGCAGAGCTAAGAGGCGATGAGAGAACCGAGAGAG 597  
QY 502 ----- 501  
Db 598 ACACAGGAAGGTGGCCAGAGAGGTGGAGACAAGGTCTGACGATGAGTGACTCTCTG 657  
QY 502 -----AGGTGTGACACATCGCTGCTTTAAAGCTGGGCTCTTCTCGGGCTCT 550  
Db 658 GACCCCGGTGCGAGGTGTGACACATCGCTGCTTTAAAGCTGGGCTCTTCTCGGGCTCT 717  
QY 551 GGTGGACCTGGCCCTGTTCTGCTGATGAGTCAAGCAGCTTCTCGAGCTTCTCGAGCTGCTG 610  
Db 718 GGTGGACCTGGCCCTGTTCTGCTGATGAGTCAAGCAGCTTCTCGAGCTTCTCGAGCTGCTG 777  
QY 611 CCTTCAACTTCCCTACCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 656  
Db 778 CCTTCAACTTCCCTACCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 823

RESULT 13  
BU234223  
LOCUS BU234223  
DEFINITION BU234223  
ACCESSION BU234223  
VERSION BU234223.1  
KEYWORDS GI:25478587  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

BU234223 698 bp mRNA linear EST 26-NOV-2002  
603792086F1 CSQCHN24 Gallus gallus cDNA clone ChEST75k21 5', mRNA  
sequence.  
BU234223  
BU234223.1 GI:25478587  
EST.  
Gallus gallus (chicken)  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;





Db 421 AATATTCCCTGATGATTCTGGGACTTCATGCACTGCGCTGCTTTGTCAGAGCTGAAG 480  
Qy 502 AGG 504  
Db 481 AGG 483

Search completed: November 21, 2004, 06:57:34  
Job time : 2289.57 secB